ON LOYALTY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Mathew Timothy Lu

August 2005

ON LOYALTY

Mathew Timothy Lu, Ph.D.

Cornell University 2005

In this dissertation, I defend a particularist, developmental account of the demands of loyalty, particularly towards unchosen objects (e.g. familial loyalties, piety, and patriotism). After surveying the limited literature in the field, I begin by pursuing Bernard Williams' suggestion that the deepest loyalties in human life are not subject to justification on the basis of universalist moral theories. In general, I argue that modern ethical theory—both broadly consequentialist and broadly Kantian—cannot account for demands of special concern towards particular persons, groups, and traditions, because those theories are intrinsically insensitive to the intra-volitional values that inform such demands.

In response to the inadequacies of the universalist views, I attempt to derive such loyalties from the intra-volitional structure of a mature human will. My argument describes the ground conditions of a morally mature will—conditions with which any tenable moral theory must be compatible. I hope to be describing some of the essential features of human moral psychology as they are actually valued, features no moral theory can ignore while remaining faithful to the content of the lived moral life.

I make special use of Harry Frankfurt's account of volition, autonomy and commitment, especially his notion of volitional necessity—the idea that a moral agent is compelled to perform certain actions not as a result of the deliberations of practical reason, but because his caring for certain objects is

itself partially constitutive of his will. However, I think this view is incomplete and so drawing on work by Jonathan Lear, I offer an account of the origins of moral responsibility in the course of personal moral development. I argue that moral responsibility can only be fully understood in light of how an agent achieves maturity as a reaction to and reflection of the public values of his social world.

I ultimately hold that the structure of any humanly valuable will is characterized by the sort of volitional necessities that give rise to the deepest demands of love and loyalty. Accordingly, any adequate conception of the mature moral agent must make room for loyalties directed at unchosen objects as acts of self expression.

BIOGRAPHICAL SKETCH

Mathew Timothy Joseph Lu was born in northeastern Tennessee in 1976 and grew up in Houston, TX, graduating from Bellaire High School in 1994. He received his A.B. in philosophy from the College of the University of Chicago in 1998 and matriculated to the Sage School of Philosophy at Cornell University in the fall of 1999. He passed his candidacy ("A") exam in June 2002, received the M.A. degree in August, and began thesis work under Richard Miller. He defended his dissertation ON LOYALTY in May of 2005 (the "B" exam), receiving the Ph.D. in August 2005.

For All My Parents

Parentibus omnibus in primis praecipientibus mihi pietatem

JMJ

ACKNOWLEDGEMENTS

I wish to thank all of the many people have helped me in my time at Cornell, including, but not limited to, the entire faculty of the Sage School of Philosophy, fellow graduate students (especially my classmate Anne Nester), and my own undergraduate students. Of all my teachers, I especially thank my committee members Scott MacDonald and Jennifer Whiting for their time, help, and comments. Above all others, I must offer the greatest appreciation to my advisor, Richard Miller, who has directed my work with sympathy, care, and never-ending consideration. Dick has taught me more than any other person, even if I have not always accepted his lessons; he has my enduring gratitude.

In the preparation and editing of this dissertation, I must thank my beloved Rachel Smith for her help in proofreading and preparing me for my defense. As well, Jacob Klein and Mark Wyman each did me a great service in reading and commenting on drafts of related material. Finally, I thank Joseph Yarbrough for help with the dedication.

I thank the Intercollegiate Studies Institute for supporting my graduate work with the Western Civilization Studies Fellowship in 2003-4. Finally, I thank the Sage School and Cornell for making my graduate work possible with the Sage Fellowship.

TABLE OF CONTENTS

When a man feels his present ties to be arbitrary or to be a mechanical bondage, he sometimes says that it is irrational to be a mere spoke in a wheel. Now, a loyal self, is always more that a spoke in a wheel. But still, at worst, it is better to be a spoke in a wheel that a spoke out of the wheel.

—Josiah Royce, *The Philosophy of Loyalty*

Part I

CHAPTER I

Loyalty: An Introduction

Where do your loyalties lie? The question has a kind of *frisson*. Loyalty is at the center of our lives, an inescapable consequence of our moral natures, and yet philosophers have not expended nearly as much effort in understanding this phenomenon as one might have expected. While one kind of loyalty (patriotism) has been picked over relentlessly, other significant loyalties—to family, institutions, traditions—have received much less attention. Indeed, the general phenomenon itself has seldom been the focus of sustained philosophical inquiry, commensurate with its place in our daily lives.

I suspect there are a variety of reasons for this state of affairs, but two are most plausible. First, loyalty *pe se* may seem straightforward and not really in need of philosophical examination. There are a variety of questions surrounding loyalty: Should one be loyal? To what should one be loyal? When should I give up my loyalties? However, loyalty itself seems simple—merely a kind of constancy not itself in need of further explanation.

This leads naturally to the second reason: loyalty *per se* seems, like other instrumental qualities, morally neutral. The value of a given loyalty seems entirely parasitic on the value of its object. Thus, a loyalty to some good end is

valuable only to the extent that the end is good, while loyalty to some evil end seems rather to intensify the evil. Like intelligence, wit, charm, courage, or any other instrumental good, the ultimate moral value of loyalty seems to lie in how it is used rather than in itself alone. As Kant points out at the beginning of his *Groundwork*, it is nature of the will that possesses such qualities that determines their moral value. For while they may all seem good *ceteris paribus*, when directed by an evil will towards an evil end all these putatively good qualities make the malefactor just that much more despicable.

However, I think this assumption is a mistake, and I think we can readily see the nature of the mistake. Unlike some of these other characteristics, loyalty is not merely a instrumental quality or tool of some wholly independent self or agency. When we reflect on the question with which I began I think we can see that to ask someone where his loyalties lie is in essence to ask the question: *who are you?* In other words, there seems to be a deep connection between an agent's loyalties and who, at the deepest level, that agent truly is. Firm loyalties reflect what the agent most deeply cares about and in so doing reveal the very ground from which his moral agency arises; or so, in the course of this dissertation I shall argue.

A Quote from Williams

In one respect, however, I have over-dramatized the lack of discussion about loyalty, because in fact there has been a significant and important debate in moral philosophy which takes up important issues closely related to the topic. In the mid-1970's Bernard Williams published an influential paper criticizing the dominant moral theories (then as now)—consequentialism and

deontological "Kantianism." I think it will be helpful to quote from its conclusion at length as I will return repeatedly to his central example and the intuitions it elicits over the course of this dissertation.

The situation[1] involves a man whose wife is trapped in a burning building with another woman, who acts to rescue his wife without subjecting his actions to moral evaluation; the question—is he justified in doing so?

> …surely *this* is a justification on behalf of the rescuer, that the person he chose to rescue was his wife? It depends on how much weight is carried by 'justification': the consideration that it was his wife is certainly, for instance, an explanation that should silence comment. But something more ambitious than this is usually intended, essentially involving the idea that moral principle can legitimate his preference, yielding the conclusion that in situations of this kind it is at least all right (morally permissible) to save one's wife…. But this construction provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife.
>
> Perhaps others will have other feelings about this case. But the point is that somewhere (and if not in this case, where?) one reaches the necessity that such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view, and that they also run the risk of offending against it.
>
> They run that risk if they exist at all; yet unless such things exist, there will not be enough substance or conviction in a man's life to compel his allegiance to life itself. Life has to have substance if anything is to have sense, including adherence to the impartial system; but if it has substance, then it cannot grant supreme importance to the impartial system, and that system's hold on it will be, at the limit, insecure. (Williams [1982], 18)

[1] The original inspiration for this sort of example comes from William Godwin's *Enquiry concerning Political Justice*, 3rd ed. (1798), vol. 1, p. 127, quoted in MacIntyre [1983], where the choice involved was between rescuing, from his burning palace, "the illustrious bishop of Cambray" versus his valet, even should the latter be "my brother, my father, or my benefactor."

This sort of challenge against the universalist (or impartial) moral theories lies at the heart of the sort of claims I wish to make on behalf of loyalty.

Loyalties interfere with neatness of the universalist moral theories because they introduce the very messy particularities of real men and women acting in the world. A man moved to rescue his wife from a fire, a mother to save her baby from danger—these situations are so natural, so important, that subjecting them to questions of moral justification seems entirely out of place—"one thought too many." Were the agents in question to have such thoughts we would really wonder whether the man truly loved his wife, or the woman fully loved her baby, in the way, deep down, we really think they should.

So under descriptions like "duties of special concern", "deep attachments", life forming "projects" and goals, the types of questions I pose in the name of loyalty have engaged moral philosophers for some time. Nonetheless, I do think it useful to gather them together and recognize that they all concern loyalty in some way. By itself, etymology is never good philosophy, but we ought to note that the word 'loyalty' is derived through French from 'law.' And that is just how loyalties seem; they are laws, but laws of the heart and of the will, internal regularities whose force often binds the agent through his will as informed by his loves, rather than through his reason as formed by his judgments.

This fact—of loyalties' law-like nature—also provides the ground for our deep intuition that loyalties are somehow essentially within the sphere of morality. I do not mean simply "morally justified" but rather that loyalties need to be considered within our larger understanding of morality, both formally within moral philosophy and in the more important informal sense

of our actual practice within the moral world. Once we see loyalties as what they really are, I think there can be little doubt as to their moral significance.

What is Loyalty?

I have already noted that loyalty has often been understood merely as a kind of constancy or steadfastness. When we speak of loyal children, friends, soldiers, etc. we generally have in mind their performing actions and having attitudes that express or uphold a commitment in the face of some sort of inducement to break that commitment. The loyal soldier is one who follows orders in the face of death; the loyal friend is one who remains firm even when doing so is costly to him in money, time, or reputation, etc. More than anything else loyalty might seem to be an attitude, a way of approaching the world that is almost pre-reflective.

It is this aspect of loyalty that often sees it ascribed to animals, particularly dogs. There is a hint of contempt in the way many approach loyalty, as if it is really only worthy of a dog. Acting from loyalty thus begins to seem more like an instinct than a reasonable course of action, and to the extent that people act from loyalty alone they can seem unreflective at best, automatons at worst. In expressing unquestioned loyalties we deliberately prescind from a certain kind of moral reflection. It is therefore wholly unsurprising that to moral philosophers—for whom recourse to extensive moral reflection is the very mark of a fully human life—loyalty will seem to be of, at best, dubious moral value.[2]

---

[2] Even steadfastness in the right—i.e. fighting the good fight—might seem best not described as a loyalty to the good, but rather a continual choice to pursue the good. In other words, to repeatedly choose the good is not quite the same thing as being loyal to the good (even if the

While there is certainly some truth in these concerns, I think the tendency to dismiss loyalty as unimportant is a mistake. This dissertation is essentially an attempt to make explicit not only the degree to which loyalty plays a significant role in human life, but also the degree to which it cannot be otherwise. Far from being only worthy of a dog, I will argue that loyalty is at the very ground of what makes us human. To see this, however, will require us to change how we approach loyalty.

My argument ultimately asserts that loyalty is a property of an authentically human-shaped will, specifically an orientation of the will that issues in actions which reflect the internal structure of that will. As a property of the will, fully human loyalties are different from instinctual animal "loyalties," and as such they represent an intrinsically *moral* phenomenon. I will argue that loyalties grow out of having a human-shaped will and are thus an inescapable facet of our moral lives. They are a reflection of the internal structure of such a human-shaped will and issue in action directly in consequence of that structure. Thus, my argument for loyalty is essentially an argument from moral psychology, which asserts that loyalties are the natural product of possessing a fully human will.

Of course central to these claims is my notion of a human-shaped will. The course of my argument will naturally embrace a larger purview than this notion alone. Most obviously I have to give some account of what I take the will to be and what would make one particular will more "human-shaped" than another. In the coming pages I will develop both of these ideas at length, but I think we can begin with a concept of the will as that (whatever it is)

---

actions performed are objectively identical in the two cases) because being loyal involves being moved to act without necessarily evaluating the nature of the obligation in each case.

which is the ground of morally responsible agency. That is, I see the will as that to which moral praise and blame is most accurately ascribed. Here I take this concept of the will very broadly, but as my argument develops it will take on more specific content. The distinction between how I see the will and how the relevantly similar capacity is described by proponents of the different strains in contemporary normative ethics will become clearer as we proceed.

To see the importance of loyalty in morality we have to examine those other strains. In particular, I want to engage, and offer criticisms of, the dominant theories of normative ethics—consequentialism and Kantianism. My criticisms of these theories will ultimately turn on a series of claims I make about the nature of value—specifically the sort of value that underwrites our moral activity. What I hope to show is how the sort of value at the heart of loyalty—and indeed much moral activity generally—is rooted in the internal structure of the will. As such, that structure (and the value it represents) will influence and even necessitate certain actions independent of an agent's practical reasoning. My criticism of the dominant theories of normative ethics is essentially that as theories of practical reasoning alone they are insensitive to the nature of the lived moral life, which is shaped by the sorts of values loyalty identifies.

To make sense of an intra-volitional theory of value I propose to take up the hierarchical mode of the will proposed by Harry Frankfurt. I think this model allows us to see how an agent can act freely (and thus be held morally responsible for his actions), while nonetheless not exercising the sort of practical reasoning most moral theories seem to assume is necessary. At the same time, however, I think Frankfurt's theories are incomplete, because they are too formalist. As a description of the *structure* of the will, Frankfurt's

theories do not give us adequate guidance as to the sort of volitional *content* we should value; they are compatible with volitional structures that should strike us as moral monstrosities.

The constraint on the theory that I intend to offer is precisely a description of what a *human-shaped* will consists in. I have a relatively naturalistic view of what is human-shaped which I derive from a picture of human development. I argue that human nature itself constrains the kinds of communities in which fully human agents can adequately develop. Some of these restrictions are derivable from an examination of how moral agents actually come to be, and I begin by examining a view of developmental psychology derived by Jonathan Lear from Freudian psychoanalysis. However, I expand the application of some of these psychoanalytical insights beyond empirical psychology and apply them more to what interests me—the will as the ground of agency.

One of the key insights of this development model concerns the relationship of the development of an individual will to the structure of the moral environment in which it comes to fruition. In particular, we must examine the relation of the developing will to the public mores and morals which, I argue, it must in some way assimilate and metabolize even to come to be at all. Of central concern in this interaction between developing will and environment is the way in which unchosen relationships are at the very foundation of the development of will. I want to argue that there is something highly significant in the fact that these relationships are, and in some important way must be, unchosen.

My argument for loyalty itself depends on my account of the development of the individual will in interaction with a moral community.

My notion of what a human-shaped will consists in lies in an account of how human nature constrains the possibilities of healthy moral communities. At the same time, I want to argue that the sorts of values that we actually do have—which inform some of our deepest moral intuitions—are the product of, and require, the kinds of structures I will describe.

The Plan of the Dissertation

This thesis is divided into two large parts, each containing four chapters. The first part is essentially negative; in it I introduce the subject, examine the pre-existing philosophical literature on loyalty, and then subject both of the two dominant strains of contemporary moral theory—consequentialism and deontological "Kantianism"—to sustained examination. The second part is the positive project, there I introduce several topics, most importantly Harry Frankfurt's hierarchical model of the will and Jonathan Lear's discussion of love in the development of the psyche. Then I attempt to achieve a synthesis of insights drawn from both, with my own development, to describe a view of the will which explains the nature of our deepest loyalties.

In **Chapter II** I examine the limited available philosophical literature on loyalty. My discussion focuses on the two philosophers who have had the most to say about the topic—Josiah Royce and Andrew Oldenquist. The late 19th century American philosopher Royce makes substantial claims for loyalty—that it is the ground of all morality—but his discussion is of limited contemporary application because he has a philosophical outlook (neo-Hegelian Idealism), which few will find appealing today. Nonetheless, he

offers insights, especially into the nature of the deep interconnection between loyalty and the self. Like anyone drawing wisdom from Hegel (however mediated) he quite properly understands, at an almost instinctual level, that to understand the nature of loyalty and morality we have to understand the nature of moral selves. Thus, we can draw considerable inspiration from Royce which points in a fruitful direction, but his arguments have to be substantially augmented to be philosophically compelling to contemporary audiences. In particular, we have to lay out and defend an account of moral psychology that plausibly fills gaps in his argument as presented.

Oldenquist's contributions to the subject are much more recent and more readily applicable. His most interesting observation is that loyalties constitute an entirely distinct category of the normative which circumscribes the scope within which universalist moral principles apply. His claim is simply that loyalty provides the non-arbitrary criterion for defining the set among whose members one is morally required to be impartial. Thus, partiality itself (in terms of loyalties) provides the delimiting factor in the moral application of impartiality. This has the implication, which Oldenquist does not fully explore, that an unrestricted impartiality offends against morality.

Finally, I close the second chapter with a brief discussion of some other philosophers' work which is not explicitly directed towards loyalty but whose application to the subject is clear upon reflection (like the passage by Williams at the beginning of this Introduction). In particular, I introduce certain criticisms of the dominant universalist moral theories which point to their defects insofar as they offend against some deep intuitions concerning the moral importance of particularity.

In **Chapter III** my intention is to assess the extent to which various consequentialist theories can accommodate the demands of special concern, and thus loyalty. I begin by defining as consequentialist any moral theory constituted by two parts—1) a theory of the good which defines the good in terms of preferred states of affairs, and 2) a theory of the right which defines the nature of the agents' obligation to promote actualization of the state of affairs defined in the theory of the good. Even from this bare structural description of the nature of consequentialist moral theories the essential nature of my objection to them is apparent. I reject, with Kant, the idea that the good can be understood in terms of states of affairs. Loyalties express particular (morally significant) values which consequentialism as such simply cannot comprehend as good.

I begin by examining the effort of Robert Goodin and Philip Pettit to derive "duties of special concern" within a consequentialist framework. They argue that such duties are consistent with a consequentialist theory of the good—which again states that only states of affairs bear independent value—by offering a "responsibility criterion" that assigns differing levels of responsibility to different agents insofar as they are individually obligated to promote the good. Thus they are duties of *special concern* because they only apply to certain, uniquely relevant individuals, but they remain *duties* within the consequentialist framework insofar as they remain directed at a universally valuable (desirable) states of affairs. I essentially argue against them that their efforts simply do not answer the challenge of loyalty, because they fail to identify the moral importance of the fact that the value of the goods aimed in loyalty lies in their particularity. In other words, they fail to offer a theory which assigns the right kind of value to the ends of special

concern, precisely because they do not understand that value in terms of the will of the agent who is bound to promote it.

I diagnose the error of Goodin and Pettit as lying not in their particular efforts but in the nature of the consequentialist project as a whole. To amplify my conclusions I consider even more sophisticated consequentialist attempts to meet the same sorts of issues developed by Peter Railton and Samuel Scheffler. Railton offers a "sophisticated consequentialism" that divorces a subjective consequentialist calculus from an objective consequentialist understanding of the good. In short, he argues that a given agent might be moved to act from motives particular to him which do not make reference to maximizing a particular state of affairs in each case, but which—taken as a whole—will have the overall objective effect of maximizing that exact state of affairs. So in given cases the agent can be subjectively moved to act without regard to the objectively preferable state of affairs, if it is the case that having such a set of motives (i.e. such a character) will in the long run most contribute to the greatest good defined by that (preferred) state of affairs. Against this I argue that this way of thinking divorces the agent's particular motivations from morality in a way that does violence to conceiving of the demands of special concern as part of a independently good will. Railton makes such motivations into non-moral personal preferences, which are allowable, but which possess no moral worth in themselves. I reject this account on the grounds that the intuitions underwriting loyalty need to be integrated into an account of the good which makes room for their place within the sorts of human will we value.

Samuel Scheffler offers perhaps the most sophisticated broadly consequentialist account, one which tries to leave room for the demands of

special concern. He offers a "hybrid" theory which maintains a consequentialist theory of the good, while offering individuals an "agent-centered prerogative" which allows them to defer maximizing the preferred state of affairs in certain circumscribed cases where doing so would alienate them from their own most important projects and goals. This prerogative allows, but does not require, this deferral; so any agent *may* always act to maximize the preferred state of affairs, but is not necessarily morally bound to do so. Against this I reiterate and expand the same sort of argument—namely, that the agent-centered prerogative does not capture the particular character of the demands of special concern insofar as they are necessary to a human-shaped good will. In particular, I again argue that Scheffler's account, attractive though it is in many ways, nonetheless still precludes the volitional nature of the good encapsulated within legitimate demands of loyalty.

In **Chapter IV** I have a similar goal but with a different focus—namely, to understand the extent to which the broadly Kantian picture of morality can accommodate the demands of special concern. In a certain, very important respect I am much more sympathetic to Kant's ethics, because I share his commitment to understanding morality entirely in terms of the good will. Where I break with Kant is in my understanding of what constitutes the good will.

I begin by offering the most sympathetic reconstruction of the Kantian moral system I can. I note that for Kant the good will is, at the limit, a *pure will*—i.e. a will determined solely by a rational respect for the moral law which itself can be partially known (as synthetic *a priori* knowledge) by means of a transcendental deduction through the categorical imperative. A pure will is moved only by motives which apply to all other rational creatures and

which may constitute universal and necessary laws. Because the good will is at the limit a pure will, Kant cannot allow the good will to be determined by particular desires, no matter how subjectively significant to a given agent. Accordingly, the demands of loyalty are intrinsically ruled out of the bounds of morality; a will moved by such demands is "heteronomous," and fundamentally unfree. Therefore, the demands of loyalty are intrinsically alien to morality.

In other words, by its very nature Kant's moral theory must be hostile to the demands of loyalty and any will which acts on such motives must be impure. An agent acting in response to the particularist demands of loyalty does not act on motives which may constitute universal and necessary laws for all rational creatures. Such motives make inescapable reference to the particular desires of particular agents. Any agent whose will is determined by his loves and loyalties is unfree, according to Kant, and so not acting as a free moral cause.

After considering Kant's moral theory, I turn to examining potentially more compelling "Kantian" theories deeply influenced by Kant, but also sensitive to some particularist concerns. Specifically, I devote an extended discussion to Barbara Herman's attempt to offer an answer the challenge of particularity, which nonetheless remains true to Kant. She argues that the particularist challenge illegitimately imports a false and misleading picture of moral psychology. Under such a picture the demands of morality and the demands of loyalty conflict with each other in a kind of psychic space into which they both flow, and whenever the demands of morality win out the agent is alienated from what he most cares about. She claims that this model offers a false choice and instead offers what she calls the "deliberative field

model" in which all demands—of morality or loyalty—are "normalized" according to the "principles of practical agency." This normalization allows them to be coherently integrated, for such principles prescribe the scope and nature of allowable deliberation. Unlike the former model, she claims this is a mature picture of moral reflection which allows the demands of special concern to be integrated into a moral life.

My criticisms of Herman are rather detailed, but in essence I dispute her claim that the principles of practical agency can actually integrate the various demands in a non-question-begging way. I argue that any such attempt inevitably undervalues the demands of special concern because it rules them out of the will and refuses to acknowledge their central place within our moral agency. Without a more nuanced understanding of the will which sees it as more complicated than pure reason alone, we cannot account for the moral value of the demands of special concern. The chapter closes with a similar examination of moral contractualism.

Chapter V inaugurates the positive project and is intended to offer a theory of the will which is sensitive to the sorts of concerns which I offer against the universalist theories. To do this I take up Harry Frankfurt's innovative hierarchical model of the will. The model holds that the will is constituted by desires of at least two levels, first-order desires whose objects are things in the world (e.g. food or drink) and higher order desires whose objects are lower order desires (e.g. the desire to desire food). Frankfurt claims that the human will is most centrally constituted by these higher-order desires. Unlike first-order desires, higher-order desires can be and are essentially reflective. As such, their satisfaction represents not merely the

satisfaction of some passing whimsy, but rather reflects what the agent, whose will they constitute, most cares about.

The signal consequence of this view most significant for my project involves what Frankfurt calls "volitional necessity." A volitional necessity involves the determination of the will to act on account of one or more of these volitionally constitutive higher order commitments. It is not the outcome of practical rationality—that is, it is not a deliberative judgment. Rather it is a direct movement of the will in response to one of its essential constituent desires. When bound by a volitional necessity the will issues in action that directly expresses its internal structure. However, such necessities manifest the agent's autonomy, because what is expressed in acting under a volitional necessity is the agent's own innermost volitional commitments and concerns. To act under a volitional necessity is to most perfectly express who I am as a moral agent, because it makes public the commitments which make up my will, my identity, my moral self.

It is these volitional necessities which provide the best model for understanding the expression of our deepest loyalties. I claim that the deepest loyalties just are volitional necessities and like them are the most authentic expression of an agent's will. The key point is that, contra Kant, such volitional necessities actually are an expression of the will's autonomy, because the necessity proceeds not from an external compulsion but the inner structure of the agent's will itself.

Despite the powerful insights I believe Frankfurt's view contains, I think there is a problem with this view that he never really addresses. Namely, Frankfurt offers a view of autonomy that makes essential reference only to the structure of the will. He does not concern himself with how the

will comes to have the structure it does. Taken purely as a structural understanding of autonomy his view is actually compatible with external compulsion insofar as that compulsion involves introducing into a victim's will higher-order commitments which serve some other agent's ends or purposes. To be a full account of volitional autonomy Frankfurt's account needs to be supplemented with an account of how an agent's centrally constitutive higher-order commitments can freely arise.

Chapter VI is offered as the start of my attempt to meet this challenge. I begin by presenting a view of the origins of the psyche offered by Jonathan Lear, which he derives from a Freudian account. My concern, however, is not with Freud, or even the precise details of the account as Lear gives it. Rather, I take Lear's account (of the etiology of individual empirical psyche) as a model for developing my own account addressed to the more interesting problem of the etiology of an agent's *moral* psyche—the ground of his moral agency.

Lear's most interesting claim is that love is the force which drives the development of the self. Love, particularly the love of a mother for her child, creates the environment in which the nascent psyche of the child actually comes to be in response to that maternal love. The maternal love creates a kind of challenge by which the nascent psyche, in struggling to meet it, is driven towards a higher and higher level of sophistication and development. At first the process is relatively straightforward, the mother's love writes itself onto the soul of the child through his identification with her, but quickly it moves beyond this as the child's soul strives to develop into the kind of thing worthy of love—viz., a *human* soul capable of loving back.

From Lear I take up the important distinction between *being held responsible* and *taking responsibility*. The former is essentially a third-personal

attitude (though one which can be taken with regards to oneself), an attitude common in straightforward moral censure or praise. To be held responsible for some act is for some external authority to declare my responsibility for it (e.g. a criminal verdict). On the other hand, taking responsibility is an essentially a first-personal act; it is in a sense a creative act, an act that involves the expression of my will into the world by claiming ownership for a part of that world. Specifically in taking moral responsibility I announce my arrival as a true moral agent, i.e. a free cause of moral action.

I argue that this model helps us to understand how not just nascent psyches but fully *moral* selves come to be. The key insight I want to draw here is that the etiology of the moral self is driven by the dialectic between the developing self and the ever more complicated loving and lovable environment in which it finds itself. It is loyalty which provides much of the driving force of this development, both as the developing soul is itself loyal to facets of the world around it (beginning with family, but quickly moving beyond to community, society, and tradition), and as the mature members of the community themselves bear loyalty to this new member and each other in sustaining a "good enough world" in which such moral development can actually take place.

It is with these considerations in mind that in **Chapter VII** I attempt to tie together Frankfurt's view of the structure of the individual will with Lear's suggestion that the individual soul is deeply dependent on the world to give it shape and form. Thus, we now have the means to answer what I took to be a defect in Frankfurt's account, his indifference to the source of an agent's higher-order commitments. We find their source in the loyalties of the nascent agent's social world, in the family and beyond.

For if this view is correct we have an understanding of where *unchosen* loyalties arise. These are the loyalties to family and community, etc. which are both the paradigmatic forms of loyalty and, under most moral theories, its the most morally suspect examples. For this account shows that some volitionally constitutive loyalties must be founded on unchosen commitments, for they are taken up from the developing agent's moral environment as the very building blocks of the self. Such foundational commitments cannot be chosen because in this context choice is meaningless; until such commitments are somehow taken up there simply is no self there to do any kind of choosing at all.

In the chapter I attempt to develop a more compelling account of the nature of the lived moral life than the dominant theories of normative ethics can offer. In particular, I try to show how loyalties lie at the center of morally responsible agency. I conclude by examining and defusing several objections to the developmental account as a whole.

Finally, in the conclusion, I seek to tie together all of these various considerations in a formal definition of loyalty in terms of an orientation of the will reflecting its internal structure as itself formed as a reflection of and in reaction to the external loving and lovable moral world. Loyalty becomes the means for understanding how moral selves are possible, and how the moral community propagates itself. Ultimately, I argue that without loyalty we can have no morality, for without loyalty we can have no moral selves. In this I hope to have given us a means to understand how even Royce's most dramatic claims for loyalty can be substantially sustained and why an understanding of loyalty has to be at the heart of any truly plausible moral theory.

CHAPTER II

A Review of the Literature

There has been surprisingly little philosophical examination of the phenomenon of loyalty as such. While particular types of loyalty—especially patriotism—have given rise to a vast secondary literature, the general phenomenon itself has been much less explored. Nevertheless, several important works do exist, and this chapter will focus on a few of them in depth.

As I suggested earlier I suspect the reason for this lack of coverage of the broader topic stems from the intuition that loyalty per se is properly regarded as a morally neutral quality. Like other traits and talents which can be put to differing uses, it might seem likely that (for purposes of moral evaluation) all that matters is the end towards which one is loyal, not the fact of one's loyalty itself. As Kant notes in the Groundwork the value of intelligence, wit, judgment, courage, resolution, perseverance, etc. lies not in themselves but in the will that expresses these qualities, and the end towards which it is directed.[1] These qualities in the service of a disordered or evil will and directed towards an evil end are far from good, but facilitate and promote evil.

[1] This discussion is towards the beginning of the *Groundwork* in and around Ak. 393.

Just as a quality like perseverance would be unlikely to spawn a large literature in itself, so loyalty has met a similar fate. On the other hand, particular loyalties, such as patriotism, bring with them implicit (or even explicit) judgments of the good. Their moral status is apparently parasitic on the moral status of the ends to which they are directed. Thus, it seems natural to subject such particular loyalties to moral and philosophical examination in a way that may seem inappropriate towards the more general phenomenon.

While this may explain the general tendency to overlook the more general phenomenon, it does not really justify it. It rests on a premise—the moral neutrality of loyalty—which is dubious. At this point, however, my goal will not be to justify the rejection of this premise, but merely to examine in detail the work of the two English-speaking philosophers—Josiah Royce and Andrew Oldenquist—who have done the most important work in this area.

Royce

The most famous book on the philosophy of loyalty is almost certainly Josiah Royce's aptly titled *The Philosophy of Loyalty*. Royce taught at Harvard in the late 19th and early 20th century and was one of the most influential of the American Idealists. The book consists of a set of eight lectures originally delivered to a non-professional audience. It is best understood in two parts. The first five lectures are dedicated to laying out what Royce himself calls "a confessedly inadequate definition of loyalty" (Royce [1995], 163). Though inadequate, it is necessary, he says, to lay the groundwork for the last two lectures in which he articulates his full understanding of the nature of loyalty.

(The sixth lecture, in which he discusses "training for loyalty," is a kind of transition between the two parts).

To the contemporary analytic reader the first part will likely prove much more congenial. The second part is clearly in line with Royce's larger Idealist project and quickly points in a direction that Royce admits some may see as a "very mystical, or at least a very fantastic thesis" (146). Indeed, in the end of his analysis, loyalty itself becomes in essence religious.

Royce is not reticent about the claims he is willing to make for loyalty. Right off the bat he asserts, "*In loyalty, where loyalty is properly defined, is the fulfillment of the whole moral law*" (9, italics in original). The proper definition, we eventually discover, is a religious understanding, but we may make some progress with his initial definition. On this initial definition loyalty is straightforwardly the "*willing and practical and thoroughgoing devotion of a person to a cause*" (Ibid.). The central consideration is his notion of a "cause." Loyalty, according to Royce, is centrally directed not to other persons or institutions, but to causes. Of course, it is possible to express loyalty to persons or institutions in Royce's system, but even this loyalty is to be understood as the commitment to a cause.

> The loyal man's cause is his cause by virtue of the assent his own will. His devotion is his own. He chooses it, or, at all events, approves it. Moreover, his devotion is a practical one. He does something. (10)

Loyalty then is an expression of will as a commitment to a cause. This volitional conception is centrally important for Royce because he understands loyalty as the central connection between an individual and his duty. In contemporary terms, Royce sees loyalty as the solution to the problem of

authority. As Royce understands it, the problem comes to the fore when as we try to understand "why my duty is my duty" when "we find [that] no external authority, viewed merely as external, can give one any reason why an act is truly right or wrong" (13). Thus, ultimately, my "duty is simply my own will brought to my clear-consciousness."

The idea here seems to be that my duty binds me through my own willing insofar as I am loyal. Your "own will and your own desire, once fully brought to self-consciousness, furnish the only valid reason for you to know what is right and good." Thus, the normative force of the moral law is inherent in my own will insofar as I properly understand that will. In other words, the problem of authority is supposed to drop away as soon as we recognize the nature of our own wills. "Loyalty, then, fixes our attention upon some cause, bids us look without ourselves to see what this unified cause is, shows us thus some one plan of action, and then says to us, 'In this cause is your life, your will, your opportunity, your fulfillment'" (21).

What he seems to imply here is that it is precisely through a loyal commitment to some "superpersonal" cause that one comes to have a self. The object of my loyal commitment is not merely external to me, but actually becomes constitutive of my will insofar as I am loyal to it. It is only insofar as I am able to commit myself loyally to some cause that I even have a will. Therefore, when I do my duty it is not as if I am acted upon by something external to my will, but rather in doing my duty I fulfill my own willing, loyal commitment to a cause. "[U]nless you can find some sort of loyalty, you cannot find unity and peace in your active living" (23).

Royce next turns to meet what he expects will be the chief complaint against his account of loyalty. Posed by the proponents of "ethical

individualism" (what we would probably call liberalism), it is basically a concern for autonomy. The complaint is that loyalty "tends to take the life out of a young man's conscience, because it makes him simply look outside of himself to see what his cause requires him to do. In other words, loyalty seems opposed to the development of that individual autonomy of the moral will which… all moralists must indeed emphasize as one of our highest goods" (31).

Does not loyalty to some external cause, runs the concern, merely involve a kind of submission to an external determination; is it not, in essence, a kind of moral cowardice insofar as one is not taking responsibility for one's own will, but rather allowing it to be determined by something outside of it? Royce's response is simple. The very objection, he says, is itself dependent on loyalty. To express this objection is to voice one's loyalty to a principle of autonomy. Thus, far from undermining the primacy of loyalty as a moral precept, it further confirms that a primal kind of loyalty is necessary for autonomy itself.

Royce makes this case by observing that "a man's self has no contents, no plans, no purposes, except those which are, in one way or another, defined for him by his social relations" (45). Since this is the case, a complete self-creation is impossible; the self is necessarily made up of content it derives from the external social world. If you don't wish to be merely an unthinking conformist, your:

> only recourse, then, is to assert your autonomy by choosing a cause, and by loyally living, and when need be, dying for that cause. Then you will not only assert yourself by your choice of a cause, but express yourself articulately by your service. The only way to be practically autonomous is to be freely loyal. (Ibid.)

Thus the very possibility of having a self is predicated on loyalty because a self comes to exist only through the "self-assertion" available in loyal commitment to a cause. It is important as well that we understand this not merely as a single choice, but expressed through temporally extended "service." In other words, it is a life (or some rather large part of a life) lived in accordance with, and expressive of, a cause.

As of yet we have only seen Royce extol loyalty in the abstract—loyalty to some cause or other. However, as he himself repeatedly insists, loyalty cannot be abstract but must be made concrete through a commitment to a real cause in the world. But now we face the obvious question—is not the quality of that to which one is loyal the determinative factor in judging of the goodness of that loyalty? In other words, is not one's loyalty only as good as the cause to which one is loyal? To be loyal to a bad cause would seem not only not to be a virtue but a positive vice. Is not the loyal criminal even more insidious than one who might be easily turned into an informant? Is not the loyal Nazi Waffen SS more morally repugnant than the German army conscript who fights only because he must to survive? In short, how can loyalty itself be inherently morally valuable when it can be put to such evil purposes?

On the other side as well, is not the loyalty to a good cause only as valuable as it is instrumentally necessary to that good cause? Is not its value completely parasitic on the value of its object? Ultimately the question becomes: why concern ourselves with loyalty at all, except insofar as it is instrumentally necessary to the furtherance of goals or causes whose intrinsic goodness is independent of any contingent factors like the loyalty of their

partisans? In other words, we might think that the good fight is good even if nobody is fighting it.

Now it seems to me that Royce need not deny this last point. Nevertheless, according to Royce, the value of loyalty is not derivative merely from the value of the cause to which it attaches. Loyalty, in itself, is valuable insofar as loyalty to practically any cause (within certain limits) is constitutive of a self. "I say, whatever their cause, the loyal express themselves" (50). This seems central to Royce's conception, because without loyalty there would be no moral agents whatsoever; so loyalty is of paramount importance in human life.

In line with this claim that loyalty is so centrally important to human life, his conclusion that the proper ultimate object of loyalty is simply loyalty itself is not surprising. Again, we need to realize that this is merely a comment about the ultimate object of loyalty, because the proximate object of loyalty for any given person will be his particular cause. Thus, actually being loyal to loyalty consists not in committing oneself to some abstract notion, but in expressing loyalty to real causes in the world. However, it is important that we see that the goodness of these causes will be judged according to their affect on the loyalties of others. "And so, a cause is good, not only for me, but for mankind, in so far as it is essentially a *loyalty to loyalty*, that is, is an aid and a furtherance of loyalty in my fellows" (56).

The idea is relatively simple. Since loyalty is itself the greatest good for humans—because it is through loyalty that moral agents come to be—so my highest duty should be to advance the cause (so to speak) of loyalty itself. However, since loyalty is only manifested through the devotion of real people

to real causes, the cause of loyalty itself is only advanced insofar as one exhibits loyalty in oneself and furthers it in others.

"In so far as it lies in your power, so choose your cause and so serve it, that, by reason of your choice and of your service, there shall be more loyalty in the world rather than less" (57). This is not an abstract principle of maximization of abstract loyalty, however, because loyalty is only made manifest in the commitment of real people to real causes, which requires each of us to choose and pursue his own cause. Similarly, it does not necessarily require that I dedicate myself to discerning which real cause, out of any that are theoretically available to me, would further the greatest number of other persons' loyalties.

I may choose my cause from among those that are (for whatever reason) "interesting, fascinating, [or] personally engrossing" (65) to me so long as it is not "destructive of loyalty in the world of my fellows" (56). Thus loyalty to loyalty does not determinatively prescribe my causes, but it does give me a way to evaluate them in terms of their permissibility (and desirability).[2] Insofar as a cause is permissible, it is a legitimate object of loyalty and one's duty of loyalty to loyalty requires that one commit oneself to this cause (whatever it may be) in a wholehearted and complete way.[3]

Loyalty to loyalty requires this wholehearted commitment to whatever cause one makes one's own because the expression of loyalty to any cause (except those inherently destructive of others' loyalties) is, in virtue of the

---

[2] What I mean by this is that the principle of loyalty to loyalty does not generate some single determinate set of duties, requiring commitment to one particular cause. Rather, it gives one the means to judge, among those causes to which one is drawn for whatever reasons, which are permissible and which are not, and even among those which are permissible which would be the best to be pursued.

[3] This does not imply that there can be no gradation in my level of commitment; however, to the sorts of life-shaping causes Royce has in mind, I must be fully committed.

social nature of causes, itself "a loyalty to my fellow's loyalty" (56). In other words, my loyalty to my cause must necessarily represent a kind of loyalty to the loyalty of those who share my cause. But the point extends even further, beyond those who are my "fellows" in virtue of their commitment to the same cause, to my fellow human beings taken as a whole. This is because (visible) loyalty to any permissible cause serves as a kind of moral exemplar to others, even those who do not share that cause.[4]

Now we are in a position to say something about some of the problem cases we considered before. Loyalty to the Nazi cause is not laudable precisely because inherent in the ideal of Nazism is the destruction of others and their loyalties. As one could not have been a loyal Nazi without the wholesale destruction of the manifold loyalties of millions of persons expressed in civil, national, and international society, so the cause of Nazism is judged inimical to "the universal loyalty of all mankind" (58). Thus, loyalty to the Nazi cause is inconsistent with a loyalty to loyalty, and as an example of the "mutually destructive conflict of loyalties," it qualifies as an example of "a supreme evil" (55).

In the end, Royce thinks all morality in general can be understood as loyalty to loyalty. He writes:

> *all those duties which we have learned to recognize as the fundamental duties of the civilized man, the duties that every man owes to every man, are to be rightly interpreted as special instances of loyalty to loyalty. In other words, all the recognized virtues can be defined in terms of our concept of loyalty.* And this is why I assert that, when

[4] Thus we are able to respect principled opposition, even if we think the opponent's cause mistaken. The consistent statesman with whom I disagree on the deepest principles of justice might earn my respect in a way that the politician who vacillates between inconsistent positions in his support of causes never does; this is so even if the former should never take up a position congenial to mine and the latter should often be in apparent practical agreement with the consequences of my position.

> rightly interpreted, loyalty is the whole duty of man. (66, italics in original).

Although he goes on to consider some additional facets of loyalty, including the nature of "conscience," we have here the main outlines of the initial, "inadequate" definition of loyalty. The "sovereign and central moral maxim" of human life is simply, "*Be loyal to loyalty*" (71). At this point, rather than merely continuing to rehearse Royce's argument, I think we will do well to begin examining it more carefully.

I fear that Royce's claim that the whole of morality is to be found in a loyalty to loyalty may strike some as cute to the point of preciousness, not to mention wildly implausible. However, I think that as we take it seriously and look at his claims with sympathy, we will find that there is more to it than what initially meets the eye.

There is a two-fold problem with his presentation: 1) these lectures were not intended "to be an elaborately technical philosophical research," but were meant to "appeal to any reader who may be fond of ideals" (xxiv). I think this leads Royce to be less precise in his articulations than he might have otherwise been and more prone to somewhat oversimplifying his thesis. 2) Royce just does not quite have all the philosophical tools (a worked out metaethics, for instance) he needs to make a fully compelling case for his theory. Nevertheless, I also think that Royce has cottoned onto an interesting, and in some deep sense correct, insight.

Let us begin with Royce's interesting claim that loyalty serves to obviate the problem of authority. Of course, I am the one who is calling it the "problem of authority." Royce talks about it as being a problem of how to bring the "external" demands of the moral law into some kind of harmony

with the internal commands of the individual will. So when we ask the questions—why does the moral law have authority over me? why should *I* do my duty?—the answer is simply because I am loyal to something (and so, ultimately, to loyalty itself).

Let us now return to this remarkable claim to which I adverted earlier: "Loyalty, then, fixes our attention upon some cause, bid us look without ourselves to see what this unified cause is, shows us thus some one plan of action, and then says to us, 'In this cause is your life, your will, your opportunity, your fulfillment'" (21). This seems to say that my will itself just is loyalty to my cause. What the cause does is unify me; it seems generative of my very self. Thus, I have to be loyal to my cause, because *that* just is what *I* am—the choice to be loyal to this particular cause. In some way my self is derived from my loyalties.

In general this might seem a very strange way to approach the matter. We might think that one has some set of interests and predilections and, based upon what one antecedently cares about, one makes choices about which causes to be loyal to. So we might think that the paradigm of loyalty is someone who looks deep within himself through a kind of sustained moral introspection and, based upon what he finds there, makes choices as to which causes he feels himself most inclined to support. It is in virtue of his understanding of the good that he chooses the causes to which he becomes loyal.

On this understanding, the self, and the will it expresses, is prior, both logically and temporally, to its particular commitments. Loyalty then is a relational property of a pre-existing self, which expresses loyalty to whatever

cause interests it. The self is clearly prior to its loyalty. Royce is opposed to this understanding.

For Royce the self cannot exist prior, either logically or temporally, to its loyal commitments, because a self just is the concrete expression of that loyalty. I am my loyalty. This seems to be clearly analogous to Harry Frankfurt's much later claims (to be examined in Chapter V) that one's central, core commitments to what one cares about are somehow partially constitutive of the will. Let us pursue this line a bit further.

A key part of what loyalty does, according to Royce, is unify the self. "[U]nless you find some sort of loyalty, you cannot find unity and peace in your active living" (23). Royce seems to think that the nascent individual will is a more or less unformed thing. "By nature I am a victim of my ancestry, a mass of world-old passions and impulses, desiring and suffering in constantly new ways as my circumstances change, and as one or another of my natural impulses comes to the front.… I have no personal will of my own" (16). Royce's solution to this problem is loyalty, which allows "a happy sort of union… between the inner and the outer, between my social world and myself, between my natural waywardness and the ways of my fellows" (19).

The idea seems to be that loyalty to some cause (whether we realize it or not) has made us into what we are—selves and moral agents. Our agency does not express our loyalties, rather it is itself generated *through* them. This comes to pass through "specific training" which focuses my internal energy and binds my natural waywardness into a coherent unity able to express itself as a will in the world. Without loyalty I become a mass of competing drives and desires, without consistency or character.

However, to begin the expression of loyalty requires "us to look without ourselves" into the social world and to copy it, at first by "mere social conformity" and "docility as an imitative creature" (19) but ultimately through a full-fledged choice to commit myself to "superpersonal" projects or ideals. Thus, for Royce the having of a self requires that that self be called forth through the loyalty of a nascent will to projects or causes which it cannot—from the beginning—fully comprehend. Rather, to some degree, it must commit itself in a kind of faith, or—as he later explicitly compares himself to William James—as "a will to believe."

If Royce is correct, then the problem of authority is an illusory one. To ask why I should do my duty is more or less equivalent to asking why I should be who I am. There is no more answer to be given than: because I must *be*. That is, because *being* (not *having*) a self requires it. Thus, in being formed through loyalty to superpersonal causes I find myself just as that creature who *is loyal*. There is no problem because the issue is not the legitimacy of external authority but the nature of the very constitution of the will. To fail to be loyal to the constitutive superpersonal cause of my choice, to fail in my most central commitments, is in some sense to lose my will (or perhaps never fully to develop one in the first place).

Thus, loyalty to loyalty is not as empty as it might have at first appeared. What loyalty to loyalty really means is the willing of morality. It is by being loyal that my self is called forth, and so by cultivating and assisting the loyalties of others I help them to call themselves forth. My loyalty to loyalty calls forth not only my own agency but helps call forth the agency of all whose loyalty I promote.

Even by just exhibiting my own loyalty in the world, I call forth the world of loyalty. That is, I make manifest the way in which the world bears a significance I both receive from and press upon it. However, as Royce makes quite explicit in his brief discussion of Nietzsche, this is not a self-created significance—it is not Will-to-Power. It is an inherited significance, which is not the less mine for being inherited. Rather, it is made my own through my loyalty to it.

Here we ought to make clear that Royce is not merely a traditionalist or social conformist. His claim is not that our selves are created by society; rather they are called forth by it.[5] Thus, keenly feeling the concerns of the "ethical individualists" Royce constantly reminds us that the will makes a choice here. We *choose* (or at least consent to) that to which we are loyal. And yet, what and who we are is partially what we become through this choice *and the service that follows it*. This is of vital importance because these partially self-constituting choices of cause are not discrete choices that, once made, stand independent, secure, and finished. Rather the choice of one's cause is just the beginning. Ultimately, we are constantly reminded, what is really important is not the choice as much as the loyalty that follows it (keeping in mind of course that the choice is constrained insofar as one cannot legitimately choose a cause that is destructive of others' loyalties). It is this loyalty, this temporally extended service made manifest through many discrete choices over the course of a life (or some significant portion of one), that gives real meaning and content to the self.

[5]Royce himself does not really speak of this phenomenon as a "calling forth," but I think this phraseology apt to describe the very complicated relation between the self and the social world into which it strived for independence whilst remaining enmeshed.

At this point we must make the transition from the initial "inadequate definition of loyalty" to Royce's final considered view. Before we begin I think it should be noted that this project only makes sense in the context of a kind of Hegelian understanding of the nature of the self. As we move forward I will make some effort to point out and partially explain some of the initially strange assumptions that seem to lay behind much of his thinking.

His movement towards his full definition of loyalty begins with a discussion of the nature of truth. He examines William James' famous statement of his Pragmatist conception of truth in order to argue that it is inadequate. I do not propose to examine Royce's claims against James partly because I am not convinced that they are entirely fair, and they are simply irrelevant to our purposes here. Ultimately Royce makes an almost Platonic claim concerning the unity of truth (i.e., that all truths, and indeed all virtues, somehow converge into something unitary). He writes:

> whoever talks of any sort of truth whatsoever, be that truth moral or scientific, the truth of common sense or the truth of a philosophy, inevitably implies, in all his assertions about truth, that the world of truth of which he speaks is a world possessing a rational and spiritual unity, is a conscious world of experience, whose type of consciousness is higher in its level than is the type of our human minds, but whose life is such that our life belongs as part to this living whole. This world of truth is the one that you must define, so I insist, if you are to regard any proposition whatever as true, and are then to tell, in a reasonable way, what you mean by the truth of that proposition.
>
> The world of truth is therefore essentially a world such as that in whose reality the loyal believe when they believe their cause to be real. Moreover, this truth world has a goodness about it, essentially like that which the loyal attribute to their causes. Truth seeking and loyalty are therefore essentially the same process of life merely viewed in two different aspects. Whoever is loyal serves what he takes to be a truth, namely, his cause. On the other hand, whoever seeks truth for its own sake fails of his business if he seeks it merely as a barren abstraction,

> that has no life in it. If a truth seeker knows his business, he is, then, in the sense of our definition, serving a cause which unifies our human life upon some higher level of spiritual being than the present human level. He is therefore essentially loyal. Truth seeking is a moral activity; and on the other hand, morality is wholly inadequate unless the light of eternal truth shines upon it. (146)

There are many interesting claims here, but we can concentrate on just a few. First, there is a claim about the unity of truth, that all actual truths somehow participate in the "world of truth." This "living whole" is somehow a "type of consciousness… higher in its level than… our human minds." It is only within the context of this unity of truth and virtue that we can understand the claim from the introduction that "[j]ustice, charity, industry, wisdom, spirituality, are all definable in terms of enlightened loyalty" (9).

Second, we have this further claim that "truth seeking and loyalty… are essentially the same process." This suggestion will be vital in helping us to understand why Royce ultimately attributes so much importance to loyalty—such as to make it "when… properly defined… the fulfillment of the whole moral law" (9).

Finally we have the claim that truth-seeking (and thus loyalty) cannot have as its object a "barren abstraction," but must be expressed in terms of real causes. This cements its connection to action—specifically, moral action. In other words, as I noted above, loyalty for Royce requires the "willing and practical and thoroughgoing devotion of a person to a cause" (Ibid). So to be loyal is to *do* something. Loyalty in the absence of real action in the world is empty; loyalty to my cause involves me making practical decisions and carrying them out. Since in the first half Royce takes himself to have shown that all justifiable loyalties are ultimately loyalty to loyalty itself, if loyalty is

indeed to be assimilated to truth-seeking, and truth is somehow unitary with all the virtues, then we come to see how practical loyalty becomes itself a species of moral action.

With this notion of eternal truth in place we can now begin to understand Royce's final and complete definition of loyalty, which comes in two related forms: 1) "*Loyalty is the will to manifest, so far as is possible, the Eternal, that is, the conscious and superhuman unity of life, in the form of the acts of an individual Self,"* and 2) "*Loyalty is the Will to Believe in something eternal, and to express that belief in the practical life of a human being*" (166, italics in original).

One remarkable thing here is the notion that loyalty somehow mediates between the self and the "Eternal," that it is through a thoroughgoing commitment to a cause that the individual self is brought into communion with the larger "world consciousness" or "world life." Thus it is loyalty to a particular, real cause which gives us the intimation of the truth which adumbrates the essential unity of the world of truth.

We should note that this is a practical effect *of our wills*. In other words, it is within our power partially to manifest the unity of Truth. This is a practical suggestion, and this allows us to understand why Royce attributes so much importance to loyalty itself, rather than the nature of any particular cause. What is important here is this notion of a "Will to Believe" (which Royce is self-consciously borrowing from William James).

In his famous essay "The Will to Belief" James writes that his thesis is: "*Our passional nature not only lawfully may, but must decide an option between propositions, whenever it is a genuine option that cannot by its nature be decided on intellectual grounds; for to say, under such circumstance, 'Do not decide, but leave the question open,' is itself a passional decision—just like deciding yes or no—and is*

attended with the same risk of losing the truth" (Castell [1948], 95, italics in original). The central idea here is that there is no non-passional way to proceed. It is not good enough to merely reserve judgment about a proposition, for that decision has the same status with respect to the will as explicitly committing oneself. There is no safe option.

In the case at hand, loyalty just is this kind of expression of will and is the only means by which the self is brought into some kind of relation with the world of truth. To care for the truth requires loyalty to some cause and ultimately loyalty to loyalty itself. For it is only by committing oneself to some real cause that one stakes out a place for oneself in the Eternal, that one fulfills one's responsibility as a Self.

For this to be remotely plausible will require one to take Royce's Idealism[6] seriously. In his metaphysics Royce seems to be a neo-Hegelian, but he combines this with a rather American, in fact almost Pragmatist, insistence on acting in the world in the service of real causes. The reason for this is quite simple. As Royce sees it, the greatest task of moral philosophy is simply to allow us to resolve the apparent contradiction between our individual autonomy and the social nature of our concrete moral laws. Moral philosophy

---

[6] The following quotations make clear the extent of his Idealism:

> The real world is therefore *not* something independent of us. It is a world whose stuff, so to speak,—whose content,—is of the nature of experience, whose structure meets, validates, and gives warrant to our active deeds, and whose whole nature is such that it can be interpreted in terms of ideas, propositions, and conscious meanings, while in turn it gives our fragmentary ideas and to our conscious life whatever connected meaning they possess. (169)

> … when I inquire about the real world, I am simply asking what contents of experience, human or superhuman, are actually and consciously found by somebody. My inquiries regarding facts, of whatever grade facts may be are therefore inevitably an effort to find out what the world's experience is. In all my common sense, then, in all my science, in all my social life, I am trying to discover what the universal conscious life which constitutes the world contains as its contents and views as its own. (171)

exists to answer the question: How can I be free if I am required to submit (to the moral law)?

The answer is now apparent. My submission to an external law, i.e., my loyalty to it, is the means by which I legitimately express my will to truth. To be compelled by the truth is not to be compelled at all; that is, it is not in any way inconsistent with autonomy. But loyalty (i.e., legitimate loyalty—loyalty to loyalty) is analogous to being compelled by the truth. In fact, since such legitimate loyalty is a loyalty to the Eternal, the world consciousness, such legitimate loyalty is itself a kind of truth-seeking.

Thus a self, insofar as it is a self, is an expression of will (i.e., a willing thing). This is a point that Hegel argues in the *Phenomenology of Spirit*. However, the will, to be a real will for a human being, to be a human self, must have content—it must express itself though a temporally extended commitment to some real cause. What is *true* is just what we find in the world-consciousness. "In so far as any of my ideas are true, my own personal ideas are therefore active processes that go on within the conscious life of the world. If my ideas are true, they succeed in agreeing with the very world consciousness that they define" (171).

Thus in making a real commitment to a real cause in some sense I presuppose the existence and necessity of the Eternal. For in embracing a cause I am, to some degree, asserting its truth. However, as we have seen, in Royce's metaphysics all that is true is part of this superpersonal world of truth. Thus my loyalty is inevitably a will to believe in the world of truth.

It should now be clear why Royce is not moved to provide an analysis of the goodness of any particular causes an agent might adopt. So long as it is a cause that is not inimical to others' loyalties—i.e., insofar as it is an

expression of a loyalty to loyalty—it will come to be part of the world consciousness. In virtue of this, however, as we maintain our loyalty to loyalty we come more and more to understand the nature of this Eternal (i.e., we come to know more of the truth), and consequently our loyalty is transformed and expanded to take the Eternal itself, insofar as this possible, as its object. This is why loyalty to a particular end ultimately engenders a universal loyalty.

We might think of it this way. Loyalty to my particular cause, which I find attractive for my own reasons, becomes a kind of admission ticket to a process that may eventually lead me towards fuller participation in the very essence of the universe. We cannot achieve direct access to the fullness of the world of truth because of our limitations, epistemological and otherwise. However, we can be loyal through nothing more than an act of the will.

In other words, though finding the Eternal is hard, being loyal—expressing loyalty to my particular cause—is relatively easy. I can choose my cause and commit myself to it without a great deal of deliberation or consternation, because it is not necessary that I choose the uniquely best cause. So long as I have even a partial understanding of the nature of loyalty what I have to do is not choose what is best, but just choose some (non-loyalty undermining) cause and fully commit myself to it. In so doing, I express my will in the world and through this expression I constitute myself (as among other things the kind of being that can enter into a relationship with—and partially constitute—the world of truth).

This happens because over time my loyalty to my cause unifies the potentially disjointed individual choices of my life into an (internally) coherent whole. Thus, what I come to be is what I have chosen to do, and my

loyalty to my cause unifies these discrete choices into a plan. In other words, it is through my loyalty to my cause that I come to have a life.

To have a life is not just to have biological existence. To have a life requires that one have a self which expresses moral agency. One thing we should see, however, is that one cannot set out to possess a life in the abstract; one only finds a life through a serious, temporally extended commitment to a particular cause. So in loyalty to my particular causes I come to have a life, and to participate in the world consciousness. This gives me my first entrée to the world of truth. It gives me, for the first time, the standing necessary to successfully gain access to the Eternal.

Thus ultimately the Eternal itself comes to be expressed in and through the loyalty of the loyal. The Eternal comes into the world through those who are loyal, by means of their loyalty. This is not experienced as external determination, but as the authentic expression of their own wills. Thus, autonomy is itself manifest through loyalty, for in my loyalty I come to be transformed. "Loyalty means a transformation of our nature" (180).[7]

At this point we might ask ourselves just how useful Royce's analysis can be if it depends on this apparently highly implausible view of the metaphysical status of truth and "the Eternal." Although I don't think Royce's

[7] We might think of it this way. In our fallen state (if we may borrow a certain metaphor for our condition) we lack access to the truth, in virtue of our lack of order. By becoming loyal to a particular cause I give some kind of form to my existence; I come to have a life. This in turn begins to give me access to the world consciousness (which, indeed, through my own willing of a cause I partially constitute). As I come to recognize this, my loyalty itself expands and begins to encompass a loyalty to loyalty itself. This can bring me to a greater awareness of the world of truth, and this will in turn become the object of my commitment. Thus eventually the world consciousness itself can become the focus of my loyalty, and in so doing I am transformed. Needless to say, we are here within the realm of theology. Royce readily acknowledges this, and conceives of his analysis of loyalty as providing an understanding of the "absolute religion."

central metaphysical commitments are that different from some of Plato's or Hegel's (or later, Heidegger's), nonetheless we might feel them out of bounds for a contemporary discussion of loyalty. One might think that any view that ultimately depends on such extremely controversial views of truth and a superpersonal world of consciousness cannot be helpful to us.

However, we are not required to accept the whole of Royce's metaphysics in order to derive something useful from his discussion. We should concentrate on this notion that loyalty to a cause is in some deep sense generative and constitutive of a self. I think Royce is right on target in thinking that what is most important in our moral philosophy is coming somehow to recognize the proper relation of the individual self to the content of a moral law that comes to us from without.

What Royce is essentially arguing is that the moral law—to the extent that we come to recognize its truth—derives its authority over us by being itself part and parcel of the way a will must be ordered even to be a self. Thus, morality does not prove to be an alien external imposition at all, but part of the very content of what it is to be free. Loyalty mediates between the self and the world by giving the former a place in the latter. Loyalty thus becomes the very core of morality by bringing us to recognize how centrally dependent each one of us is on the social world which we inhabit.

Furthermore, loyalty is a practical demand, and one it is easy to recognize the value of. We each of us are almost instinctively loyal. It requires little effort then to come to express some kind of loyalty. This becomes the first, necessary step that eventuates in a moral self. This is an important insight, but it seems to me that to be most useful for our purposes it requires less that we discern the Eternal than that we come to explain how the actual

contents of our loyalties are important to the kinds of selves we have reason to prefer. And to do this, I submit, will require us to pay attention to the way in which loyalty is central in the very development of a self.

At this point I mainly wish to emphasize this notion that loyalty mediates between the self and the world. Without concerning ourselves with the "Eternal" and the "world of truth" we can still ask ourselves whether these apparently outlandish claims nonetheless engage some of our intuitions, even peripherally. I believe they do, and when these claims are augmented by a much more substantial and worked out hierarchical model of the will, supplemented by a developmental account of moral maturity, I think this will be much more apparent. However, that must await Part II

Oldenquist

Having considered the most famous book on the loyalty—and derived at least some inspiration from it—I want now to turn the contemporary philosopher who has made perhaps the most important contribution to the explicitly philosophical study of loyalty. In his essay "Loyalties" Andrew Oldenquist argues that "Anglo-American philosophy has ignored an important area of the normative" (Oldenquist [1982], 173). That essay is dedicated to capturing that overlooked terrain by sympathetically focusing on the importance of loyalties in human life.

His single most interesting claim is easy to identify: "In terms of the logic of the reasons they provide, loyalties are a third category of the normative, distinct from both self-interest and impersonal morality" (176). We should be clear that this is a claim about "the logic of reasons"—i.e., the forms

reasons can legitimately take. The first two "categories of the normative" are familiar to moral philosophy.

When giving reasons for his actions an agent's appealing to his self-interest is readily understandable. If asked why he did X, the rational egoist could simply reply "Because I judged that doing X would be good for me" and we would understand his rationale completely. This is not to say that we would approve of it, or even agree with his judgment (i.e., we might think he is mistaken in thinking that doing X was good for him). The point is merely that self-interest is a recognizable norm by which to order one's actions, an appeal to which is a rational (if weak) response to a demand for the giving of reasons for why one acted in some way.

Similarly, "impersonal morality" also constitutes a norm that can ground reasons. By "ground reasons" I simply mean that these categories of the normative can be self-sufficient stopping points in the regressive appeal to reasons. "Because (I think) it is good for me" or "because (I think) impersonal morality requires it" are both adequate answers to "why did you do that?"—adequate not necessarily as full justifications (though they may be), but as reasons of the sort that can stand on their own. This is by no means to imply that reasons of this sort need end our investigation; there may continue to be questions of whether the agent is correct in his judgments and questions about whether further judgments about the content of morality are necessary. Nonetheless, for many philosophers reasons from either or both of these two categories have seemed intuitively plausible as grounding reasons.

Thus Oldenquist is actually making a rather remarkable suggestion in claiming that loyalties constitute a truly separate "category of the normative." One might wonder whether "because my loyalty required it" is an adequate

reason in the way that our previous "because" clauses were. One might further wonder whether or not an appeal to loyalty could not be reduced to one of the other categories. Thus, one might suppose that "because my loyalty required it" should be analyzed as "because (I think) doing what loyalty requires is good for me" or as "because (I think) doing what loyalty requires is itself required by impersonal morality—i.e., I have a impersonal duty to do what loyalty requires." If this were the case then loyalties would not constitute a true third category of the normative because they could be reduced to one or both the other two.

Oldenquist claims that loyalties "cannot be understood if we try to turn them into either impersonal duties or sophisticated egoism." They cannot be impersonal duties "since an obligation of loyalty depends on viewing a thing as one's own." They cannot be mere appeals to self-interest "because people can sacrifice, in the name of loyalty, their happiness and even their lives, and it is probably this element of self-sacrifice that makes most people classify motives of loyalty as moral motives. Moreover, reasons of loyalty have a general appeal among members of a society whereas a self interest reason appeals only to the agent." The second consideration is more plausible than the first because it does not seem that odd that there might be an appeal to personal happiness that goes beyond immediate pleasure or even existence in this world (e.g., a crude kind of afterlife reward of happiness).

It might be objected that even if two people share a reason for loyalty to a shared interest—e.g., they are both members of the same family—in fact the reason can be reduced, for each of them, to "because doing what loyalty requires is good for me (all things considered, in the long run)." It may be true that both people might think that way, but having such a reason is really very

different than "because loyalty requires doing this." In other words, if they are making the reduction then they are simply making a mistake about loyalty—they have not understood that loyalty can constitute a reason at all.

Consider that straightforward moral duties might be reduced in the same way—e.g., I did not lie because doing this is good for me (all things considered, in the long run). The mere possibility of this linguistic reduction does not, by itself, show that morality can be reduced to enlightened egoism, precisely because we tend to recognize that there is something essentially non-self-regarding in reasons of impersonal morality. Similarly, there is something essentially non-self-regarding in reasons of loyalty.

This observation, however, might make us question whether "most people" are not right in thinking that reasons of loyalty are a species of moral reasons. Oldenquist argues that the essential structure of loyalties—that they only attach to that of which I have a kind of ownership (as a member of a group)—precludes their being a species of impersonal moral reasons. These latter, we might think, attach in virtue of some intrinsic properties of agents and patients, not to their relational properties. Loyalties make essential and unmistakable reference to what William Godwin called (disparagingly) the "magic… in the pronoun 'my.'"[8] We might consider Oldenquist's project to be an attempt to demystify to some extent what exactly that magic consists in, and thus make it perhaps seem less magical in the negative sense (i.e., mystical) and yet more magical in the positive sense (i.e., beguiling).

Having asserted the existence of this third "category of the normative" Oldenquist wants to get some mileage out of it by claiming: "Our wide and

[8] MacIntyre quotes this in "The Magic in the Pronoun 'My'" (MacIntyre [1983]),p. 122; the original source is William Godwin, *Enquiry concerning Political Justice*, 3rd ed. (1798), vol 1, p. 127.

narrow loyalties define moral communities or domains within which we are willing to universalize moral judgments, treat equals equally, protect the common good, and in other ways adopt the familiar machinery of impersonal morality" (177). In other words, it is loyalty which marks out the scope within which we will apply impartial judgments. He gives an example regarding the provision of resources between children within a family and the neighbor children. Between two brothers it seems correct that their parents should treat them (approximately) equally in giving them gifts, and no wrong is done if their parents do not include the neighbor child in the largess.[9] Similarly, my loyalty to my fellow countrymen (including our shared loyalty to our country) would justify some welfare benefits for their sakes, even though, objectively speaking, much more good could be done for much poorer people in the Third World.

Oldenquist notes that modern man is enmeshed simultaneously in many "nested" "'tribes' or moral communities" (177). Like "concentric circles" many of my communities are themselves part of broader communities of which I am also a member. Thus, my family is part of my neighborhood, which is part of my town, which is part of my state, which is part of my country, etc. The widest loyalty communities may encompass our entire species (or perhaps even more—e.g., all suffering creatures or perhaps our fellow members of the Federation of Planets, etc.). Within each of the communities (or domains) I have reason to be impartial, but between each of these domains I may have reason to be quite partial. Thus, I may favor the

[9] It should be obvious, but this does not mean that in all cases I *must* prefer my children to the neighbor child. If it came down to giving my kids a new toy or preventing the neighbor child from starving, it seems plausible to think that I would do wrong not to discharge my resources for the neighbor child's benefit over that of my own children. Nonetheless, the principle should hold in most cases.

interests of my family over the interests of my town, etc. However, there need be no hard and fast rule that says I must favor an inner circle over an outer one; circumstances, including what is at stake and the level of harm, etc. will determine which direction my loyalties will pull most.

Oldenquist pushes this even further and claims that without loyalty impartial morality itself cannot make complete sense. This is because an impartial moral view must give some non-arbitrary account of who counts in our moral deliberation and why. So, for example, the utilitarian has to give some reason for why non-human suffering would not count as much as human suffering in the calculus of pleasure.[10] The Kantian would have to explain why non-rational beings are not part of the moral community.[11] Oldenquist thinks that group loyalties provide the most plausible explanation of who counts and why. "On this view all morality is tribal morality and there will be as many systems of social morality as there are loyalties" (179).[12]

> …We are often told to look beyond our neighborhood, city, or country in the name of impartiality, and treat what we had hitherto favored as our own as just one among many

[10] I duly note that some "animal rights" type consequentialists bite this bullet, but most seem to hew toward Mill's plausible (but undefended) assertion that it is better to be a Socrates dissatisfied than a pig satisfied. See Chapter III for an extended discussion of consequentialism.

[11] I am here reporting Oldenquist's challenge, not endorsing it. I think that the Kantian would have a ready answer to this question, but it would turn on Kant's claim that moral knowledge is "synthetic a priori knowledge" that can only be had through a transcendental deduction. See Chapter IV for an extended discussion of Kantian moral philosophy.

[12] Alasdair MacIntyre seems to make a similar point in a discussion of one of Bernard Williams' criticisms of the way impartial moral theories inevitably alienate agents from their "deep attachments." He notes that this "kind of problem only arises… for a variety of utilitarian and Kantian writers, because they liquidate the requirements of justice within the household or family, the requirements of justice within the political community, and the requirements of justice in a variety of other particular spheres (church and school, for instance) into a conception of the requirements of justice, as such, the imposition of which makes almost all social particularity irrelevant. Thus it is not impersonality and impartiality as such that create those problems…. It is rather an impersonality and an impartiality required to hold between all persons whatsoever equally, a socially contextless impersonality and impartiality which is the source of those problems" (MacIntyre [1983], 123).

> neighborhoods, cities, or countries. But the demand for impartiality is never true impartiality, it is merely an invitation to give one's loyalty to a larger whole with which someone identifies; in other words, an invitation to join a larger tribe. If our first love is to some narrower group, this forced shift may render our moral concern weak and pallid. Equal moral concern for the whole of humanity or the whole of sentient nature is, for most of us, too diluted to be able to generate effective moral enthusiasm and too weak to outweigh narrower loyalties. If one thinks that that wider loyalty must carry greater weight just because it is wider, it is important to remember that it identifies one's kind under a different description than the narrower loyalty. The nature of the description as well as the size of the domain it encompasses determine how much I care about my kind under that description. This consideration by itself refutes the utilitarian claim that more is always better. I have extremely weak galaxy-member loyalty relative to species loyalty, and weak species loyalty relative to family loyalty. More is not always better because it not always more of the same thing. Hence, without further argument it will not do to claim, in the name of speciesism, that nationalism, community loyalty, and family loyalty count for less because the count for fewer. (181-2)

The key claim here is this idea that "the demand for impartiality is never true impartiality" for the scope within which impartiality applies can only be determined by another, wider loyalty. Even when we think ourselves impartial between all members of the human race, we still are implicitly expressing a norm which is grounded in a (species-wide) loyalty. Thus, the claims of impartial morality to automatically trump the demands of loyalty are really best understood, claims Oldenquist, as the demand of a wider loyalty to trump a narrower one. He seems to think that hiding behind a claim of impartial morality is no escape from dealing with loyalty, because some form of loyalty is implicitly operating in whatever way one defines as the scope of that impartiality.

What this suggests of course is that there is no such thing as impartial morality, simpliciter, for Oldenquist. Instead, since loyalty constrains the

scope within which impartial judgments are made, loyalty is always more basic than "impartial morality." Impartial morality is, at best, impartial judgment within a very wide scope loyalty.

We need now to get a better handle on just what sort of reasons loyalty supplies for Oldenquist. When loyalty is providing a ground reason for my action, what matters is that I am attached to the object by some bond that makes essential reference to that particular; the bond holds not in virtue of properties which that object could in principle share with a qualitative duplicate, but in virtue of its relationship to me as the one particular thing to which I am attached.

Oldenquist considers an example in which a father, confronted by the situation in which both his daughter and her friend are drowning (and he is unable to save both), acts to save his daughter. It is not in virtue of impartial judgments about his daughter's superiority that causes him to save her rather than her friend. Even if, objectively, the friend is superior (in intelligence or even virtue, perhaps), the father acts out of loyalty to his daughter simply because she is *his* daughter. Furthermore, if the friend were somehow qualitatively identical to his daughter (except for the relationship) this would change nothing. The qualitative duplicate could never (for example) substitute for his daughter (though he might come to love the duplicate in her own right if he adopted her, perhaps). Nonetheless, it is to this particular girl (his daughter) that reasons grounded in his loyalty must make reference. He might also bear some sort of loyalty to other of his children, or even his daughter's friend, but those loyalties will need to be explained in virtue of those particulars.

Any attempt to give reasons for his actions that reduce them to a species of impartial moral reasons must inevitably miss the point of the attachment and the way it underwrites his loyalty. If he acts to save his daughter not because she is his daughter but because he makes a judgment that this situation instantiates some more general moral principle (e.g., I am morally obligated to save drowning children, ceteris paribus), then he has not acted on a principle of loyalty. However, it seems we have reason to think that this kind of reasoning—while certainly possible—is not what animates the father in that situation, and, if it were, would be somehow inappropriate.

Consider the case if the father were a consequentialist and he (correctly) judges that saving the friend would better maximize the principle of the good he accepts and so ignores his daughter for the friend. In such a case, notes Oldenquist, we would feel contempt for a man who would let an abstract moral judgment overwhelm a loyalty to his own children.

> It is not just that people are as a matter of fact blindly loyal or biased toward various social units, this being a fact of human nature that we lament; rather, loyalty behavior elicits approbation and opposite behavior typically elicits guilt in the agent and disapproval in observers. The contempt we feel toward traitors is not unlike what we fell toward the father who lets his daughter drawn. Our loyalties are values that contribute to what we think we should do, all things considered. Therefore that can compete with what are called "considerations of social morality." This is even more plausible if, as was suggested earlier, every group loyalty creates and grounds a domain of social morality. When these loyalties are themselves challenged in the name of "social morality," the wider obligations determined by a wider loyalty are being asserted to have greater moral authority. As in any case of conflicting normative considerations, sometimes a person will judge his family or national obligations to take precedence over wider societal obligations and sometimes he will not. It depends, among other things, on how much is at stake in each domain, on the

> possibility of a given action satisfying both loyalties to differing degrees, and on the "strengths" of the loyalties themselves. (187)

This summary of his argument points to what may be its most significant weakness. Oldenquist adverts repeatedly to how we would feel in observing or participating in various situations where loyalties are either expressed or denied. One imagines that the impartial moralist would object that such feelings—especially feelings like contempt—are simply irrelevant to the moral evaluation of the actions in question. It may be true, psychologically, that we would feel just the way that Oldenquist describes, but of course that does nothing to show that we ought to feel that way, or that those feelings have any moral significance whatsoever.

This objection does not require the impartial moralist to give no weight to moral feeling at all (though he may be inclined to do just that). In any case, whether or not feelings have a place in a moral theory, the sorts of situations that Oldenquist discusses, and the intuitions to which he appeals, are not sufficient by themselves to prove his point. He would need to give a moral argument for the moral applicability of the feelings. The mere fact of their existence no more excuses a breach of impartial morality than the kleptomaniac's strong desire to steal justifies his thievery.

In short, Oldenquist has far from proven his point, because the putative intuitions to which he points may themselves be subject to moral criticism. It may be that we simply ought not to have such intuitions, and that in a more perfect, more just world we simply would not. Perhaps the strong intuitions underlying the feelings in support of loyalty are simply the product of contingent social circumstances that are subject to alteration and indeed ought to be altered in light of an impartial ethic of justice.

Oldenquist's response to this sort of objection can be derived from his book, the Non-Suicidal Society, which he developed out of the essay we have been examining. That response is relatively straightforward, but moves in a direction I suspect many moral philosophers would find deeply unsatisfying. In that book Oldenquist argues a relatively simple claim to justify loyalties. He thinks that human nature, as developed in the environment of selection in which our genetic forbearers supposedly arose, has most suited us—psychologically—for relatively small group loyalties.

He notes that one of the most significant problems facing modern society is the frustration of our human (and presumably genetically determined) need for relatively small human community, on account of which we become alienated from our actual societies. This alienation leads to moral corruption, as people become isolated and disaffected with the world around them. Thus, he essentially argues that to resurrect our ailing society (hence the title) we need to focus on recreating opportunities for the expression of small group loyalties within modern society. Substitute objects of loyalties will model those from our evolutionarily significant past and so satisfy our human needs for community and belonging.

We can see that Oldenquist would respond to the impartial moralist's charges with a series of claims about human nature. Drawing on evolutionary biology, he makes claims predicated on the thesis that modern human beings evolved in small communities and so are still psychologically best suited to social orders where small group loyalties can be expressed. Accordingly, membership in such groups is required for happiness and for agents' investment in social norms and institutions. Thus, he would claim that the feelings and intuitions to which the argument for the normative status of

loyalty appeals supervene on facts about human nature. Given the relative fixity of that human nature, he can then claim that any moral theory that requires the frustration of these psychological properties will be, quite literally, inhuman.

Even from this short précis one can get the sense that Oldenquist's book is an unusual one for a philosopher and may in fact seem more like a piece of (at best) philosophical anthropology than moral philosophy. I suspect that the impartial moralist will simply be unmoved by this analysis, for it is always open to him to continue to insist that these facts (if facts they be) about human nature cannot by themselves generate a normative argument for loyalty. It is always open to the impartial moralist to argue that justice requires we get over our psychological attachment to small group loyalties, even if we happen to be psychologically best suited to them. This sort of argument from psychology simply cannot constitute an adequate *moral* argument.

Here we have actually hit a kind of impasse, the significance of which I hope will become more clear in the chapters that follow. For I too will be making a kind of argument about the how loyalties are integral to human nature. However, I will take a much different tack than Oldenquist, as my argument will focus on the nature and constitution of the will. As such, I have hopes that my argument may be more persuasive to the impartial moralist than Oldenquist's can be, because my appeal will be to a faculty—the will—which is much more evidently the proper subject of moral philosophy and moral psychology than the sort of almost empirical claims which Oldenquist advances.

That said, I have deep sympathies for Oldenquist's project and conclusions. However, I think that it is possible to give an argument better

focused against the impartial moralist than the sort of naturalist argument that Oldenquist adopts. Thus, even if Oldenquist is correct in his claims about human nature (I have my doubts but will reserve judgment here), his moral argument can only be sustained to the extent that the necessity of loyalty as integral to the moral agent can be established. In short, Oldenquist's argument becomes too psychological and empirical to be fully philosophically satisfying. It must be supplemented with an argument whose focus is on the will—the locus of moral activity—and not the empirical psychology of human beings. Only then will his conclusions—many of which I find intuitively plausible—be vindicated in a manner that at least meets the impartial moralist on his own ground.

Beyond Loyalty, *per se*

It is certainly the case that other philosophers have also considered the normative status of loyalty. However, I do think that Royce and Oldenquist are the most significant for my purposes here. Therefore, in what follows I will touch on a few developments in contemporary moral philosophy that do not focus on loyalty *per se* but which are important for understanding where my project fits in the larger debates of our time. However, I will not be delving into the details of these arguments as I did for Royce and Oldenquist. The reason for this is simple: I will cover the most important of these (for my purposes) in the next two chapters when I look first at what Consequentialists have to say about loyalties (Chapter III), and then how broadly Kantian views consider the matter (Chapter IV). Thus, my purpose in the last section of this chapter is mainly to introduce some of the background to the debates in moral

philosophy into which I am wading—*in media res*, and for my own slightly idiosyncratic purposes—in the chapters which follow.

Perhaps the most important pre-existing debate from which I am drawing materials is the continuing controversy over the whether the most influential contemporary moral theories—consequentialism and Kantianism—can be reconciled with the felt important of certain intuitions attaching to what Bernard Williams called our "deep attachments." In particular, the debate has come to a head over whether such moral theories are so universalist with respect to their basic premises that they simply can never truly see particular attachments as morally significant.

Bernard Williams has played a very important role in opening this line of debate. This strain in his thinking came to the fore in his seminal critique of utilitarianism, which has come to be called the objection from personal integrity. In the context of a debate with J. J. C. Smart, Williams wrote, "how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because someone else's projects have so structured the causal scene that that is how the utilitarian sum comes out?" This would be "to alienate him in a real sense from his actions and the source of his action in his own convictions" (Smart and Williams [1973], 116).

This objection basically turns on the fact that the utilitarian theory makes no distinctions concerning the levels of responsibility which govern when and where the moral agent is required to act. Thus, the fact that my own convictions, plans, projects are *mine* is irrelevant in determining what I am to do when faced with a moral choice. I must always act in such a way as to maximize the good, even if that requires me to drop everything that matters to

me and respond to the actions of some person wholly unconnected with me. Such a requirement, argues Williams, is entirely unreasonable for it would "neglect the extent to which *his* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which [the agent] is most closely identified" and so would constitute "in the most literal sense, an attack on his integrity."

I do not wish to go deeply into this issue here because it will be taken up again in the next chapter. I merely want to introduce the nature of the objection. Williams continued to develop this type of objection in the years that followed. A particularly important statement of them can be found in his essay "Persons, Character, and Morality" from which I have already quoted in the Introduction and to which I will again return in the following chapters. There he continued to press the notion that modern ethical theories (both consequentialist and broadly Kantian) are in some ways inevitably alienating. They force a moral agent to take up an attitude towards himself and others which is inconsistent with the expression of "such things as deep attachments to other persons." Yet Williams thinks such attachments are absolutely necessary, for without them "there will not be enough substance or conviction in a man's life to compel his allegiance to life itself" (Williams [1976], 18).

Williams continued in a similar line in his book *Ethics and the Limits of Philosophy*, which ultimately seems to indict philosophy as being more of a hindrance than a help in the moral life. He argues there that our ethical theories are overly reductionist, that they attempt to describe the "truth about… the ethical" with only "one or two ethical concepts, such as *duty* or *good state of affairs*" (Williams [1985], 17). Ultimately he argues "that philosophy should not try to produce ethical theory", and "that in ethics the

reductive enterprise has no justification and should disappear." I cannot begin to examine the scope of this argument (though I will touch on very similar themes from Williams and others in the following chapters, especially in Chapter VII). Nonetheless, in this book Williams launched a kind of war against the entire project of philosophical ethics as it has been practiced to that time (and continues to be practiced to this day). Whether or not we are sympathetic to his rather strict claim about the "limits of philosophy" in justifying an ethical life, we must nonetheless recognize how important Williams has been in pointing out how much divergence there really is between most ethical theories and the real substance of a lived, ethical life.

About the same time as Williams was formulating his objections to utilitarianism, Michael Stocker published "The Schizophrenia of Modern Ethical Theories" in which he argued that such theories, by dealing "only with reasons, with values, with what justifies," fail "to examine motives and the motivational structures and constraints of ethical life" and in virtue of this "fail as ethical theories" (Stocker [1976], 453). The schizophrenia consists in the fact that theories do not adequately value the *person*, "not merely the person's general values nor even the person-qua-producer-or-possessor-of-general-values" (459).

The defect Stocker is identifying lies in the fact that such theories have the agent focusing on the wrong thing. "[L]ove, friendship, affection, fellow feeling, and community all require that the other person be an essential part of what is valued." In other words, it is the particular individual, not the other as an instantiation of some broader principle which must be central to our ethics. It is the other herself who is valuable, not the fact that she is intelligent or beautiful or kind, etc. Stocker think that contemporary ethical theories only

focus on such "general values" and thus construe justification entirely in non-particular terms.

In short, Stocker's claim is that "if we take as motives, embody in our motives, those various things which recent ethical theories hold to be ultimately good or right, we will, of necessity, be unable to have…" the sorts of motives necessary "to realize the great goods of love, friendship, affection, fellow feeling, and community" (461). The key idea is that such ethical theories (both "utilitarianisms" and "deontologies") simply rule out, in their descriptions of morally acceptable motivations, those motives which necessarily underwrite the goods of love and friendship, etc.

Thus Stocker's argument, as many of these arguments are, is a kind of *reductio ad absurdum*. He attempts to show that if you take seriously what both utilitarian and deontological ethical theorists have to say about allowable motives, you then rule out the particularist motives that undergird the true valuing of people and selves—including, he notes, our own selves. This, presumably, is unacceptable as it contradicts our deep intuition that goods like love and friendship are indeed among the greatest of goods. We cannot take the theories seriously and still hold onto what Stocker takes to be some of the most essential motivations in our moral lives. The "schizophrenia" lies the way in which such theories divide us from ourselves and others "in regard to the personal relationships of love, friendship, affection, fellow feeling, and community."

A third significant philosopher to express doubts along similar lines about the nature of modern ethical theories is Alasdair MacIntyre. In his 1981 book After Virtue, he presented a picture of the emptiness of contemporary moral theory that was nothing less than shocking in its audacity. His central

claim is: in "the actual world we inhabit" what we possess of the language of morality are merely "the fragments of a conceptual scheme, parts which now lack those contexts from which their significance is derived. We possess indeed simulacra of morality, we continue to use many of the key expressions… [but] we have—very largely, if not entirely—lost our comprehension, both theoretical and practical, [of] morality" (MacIntyre [1984a], 2).

MacIntyre's argument is essentially that our moral language is merely left over from a now destroyed Aristotelian virtue ethics past. It was that virtue ethics, and particularly its teleological conception of human nature and the good, which provided the "contexts from which [the] significance" of moral concepts was derived (or derivable). Without such a context—and he makes the historical claim that it has been lost—our use of moral terms, concepts, and categories is empty.

Obviously, we cannot here examine this remarkable thesis. What I mostly want to point out is MacIntyre's sympathy with the general line of criticism of the (universalist) modern moral theories. This is especially clear in MacIntyre's 1984 Lindley Lecture "Is Patriotism a Virtue?" Though MacIntyre rather scrupulously avoids actually answering his own question, he does, however, set up a dichotomy between two entirely incompatible and incommensurable "moralities"—the morality of patriotism and the morality of modern liberalism. This again basically reflects the division of ethics—and presumably intuitions about ethics—between those who favor a particularist (patriotic) morality and those who favor a universalist (liberal) morality.

The morality of patriotism is reflected in the following claims:

> Detached from my community, I will be apt to lose my hold upon all genuine standards of judgment. Loyalty to that community, to the hierarchy of particular kinship, particular local community and particular natural community, is… [thus] a prerequisite for morality. So patriotism and those loyalties cognate to it are not just virtues but central virtues. (MacIntyre [1984b], 11)

The central idea in the morality of patriotism is that the ethical life is fundamentally inseparable from the loyalty that binds the individual to his community. It is not merely that, all things considered, loyalty will conduce to the best moral outcome, but the much stronger claim that without loyalties like patriotism morality itself is impossible. These loyalties are fundamentally intrinsic to ethics.

MacIntyre contrasts this with the morality of liberalism which is characterized by a entirely different conception of the moral.

> first, that morality is constituted by rules to which any rational person would under ideal conditions give assent; secondly, that those rules impose constrains upon and are neutral between rival and competing interests—morality itself is not the expression of any particular interest; thirdly, that those rules are also neutral between rival and competing sets of beliefs about what the best way for human beings to live is; fourthly, that the units which provide the subject-matter of morality as well as its agents are individual human beings and that in moral evaluation each individual is to count for one and nobody for more than one; and fifthly, that the standpoint of the moral agent constituted by allegiance to these rules is one and the same for all moral agents and as such is independent of all social particularity. (MacIntyre [1984b], 7-8)

The key idea here is also relatively simple: ethics is to be understood as completely divorced from "social particularity." What matters is an objective description of the nature of the human individual (moral agent), and it is with respect to the individual alone that any moral rights and duties are to be

understood. Such a view is putatively neutral before competing "special conception of the good" (as Rawls called them), for it is not based on any explicit claims about the purpose or end of human nature. Rather it is founded on a conception of "ideal conditions" in which "reasonable" agents give their consent to rules of conduct and social intercourse (at least at some basic level).

As I noted before, MacIntyre does not try to settle the debate between these competing moralities, each of which—in his view—suffers defects. Ultimately we come to see that so barely described, each of the competing alternatives is inadequate. In After Virtue, MacIntyre finally holds that—consistent with his thesis I noted above—neither of the proposed alternatives truly represents a recovery of the lost teleological context which he sees as necessary for a genuine ethics. That said, I think he is generally more receptive to the claims of the morality of patriotism, but they need to be modified to be consistent with the "tradition of the virtues" and the shared striving for the good in a moral community committed, as a whole, to that good.

The key modification is the reintroduction of a shared conception of the good for man based on an understanding of virtue. A true patriotism is only possible in a society possessed of a government that expresses or represents "the moral community of its citizens" (MacIntyre [1984a], 254). Lacking such a shared commitment today our modern societies "lack in the fullest sense a patria" and so a genuine patriotism is simply impossible. "Modern systematic politics, whether liberal, conservative, radical, or socialist, simply has to be rejected from a standpoint that owes genuine allegiance to the tradition of the virtues; for modern politics itself expresses in its institutional forms a systematic rejection of that tradition" (255). Thus, "loyalty to my country, to

my community—which remains unalterably a central virtue—becomes detached from obedience to the government which happens to rule me."

My goal in this very quick examination of MacIntyre has not been to explore fully his arguments for their great intrinsic philosophical interest. Rather, what I hope I have done is show how three particularly influential philosophers have fundamentally altered the contemporary debate over both the status and nature of moral philosophy itself and its relationship to ethics as lived. I have merely wanted to set up some of the apparatus surrounding the debate, and though I assume it is obvious where my sympathies lie, nonetheless I have as of yet not even begun the presentation of my argument.

In short, my purpose in this last section of this chapter has not been to advance a particular argument for the superiority of the particular over the universal. Rather, I have only hoped to show that the debates into which I am offering my argument for loyalty are not as settled as the main practitioners of the dominant traditions might assume. For the debate is not properly the one so often assumed, between consequentialism and various liberal deontologies, but between either such type of universalist moral theory and the actual practice of the ethic life informed by the particular goods of friendship, family, community, country, and religion which seem intrinsically particular.

The overall debate has now become shaped into a disagreement over the essential universality or particularity of ethics itself. This is, of course, intrinsically related to a continuing debate between advocates of cosmopolitan and communitarian conceptions of community and justice. I do not propose to expand further on the background literature because I think we should now be reasonably clear on the nature of one of the most important sub-debates

into the heart of which this dissertation will be aimed. There are other cognate and not so cognate literatures which will have an important bearing on my argument—particularly in moral psychology—but I will introduce them in the course of my argument in following chapters so that I can expose them, unlike the précis of this chapter, to more sustained philosophical examination. The next chapter begins this examination with a discussion of consequentialism.

CHAPTER III

Consequentialism and the "Demands of Special Concern"

In the previous chapters I attempted to motivate the prima facie challenge loyalty poses to the most prevalent theories of normative ethics. My contention is that common sense morality gives an important place to the concerns of loyalty—including and especially towards particular, unchosen objects—which is incompatible with the largely universalist nature of those theories. I take these intuitions about loyalty to be morally important phenomena, worthy of careful attention; however, as of yet I have not presented a positive case for taking the demands of loyalty to be legitimately binding.

It is possible that the particularist demands of loyalty are simply unjustified and that the clear sense many people have that they constitute valid demands is merely the mistaken prejudice of sentiment that many critics have claimed it to be. My defense of this particularist conception of the demands of loyalty must wait for Part II. Nonetheless, there are important middle positions as well, and indeed I think most modern moral theorists occupy one of these.

These theorists acknowledge that there is some real moral content to the conceptions of common sense morality regarding the demands of loyalty.

However, they hold that this real content is in some way derivative from more basic universal moral commitments. They would think that the great priority common sense morality seems to give to loyalty is in fact a mistake—a relatively innocent mistake perhaps, but a mistake nonetheless. The goal of the moral theorist then becomes to show how any genuine moral content in the demands of common sense morality can be accounted for within the confines of his moral theory. The remainder is either rejected outright or, more likely, demoted in status to become a kind of permissible (personal) preference rather than a valid obligation.

In the present chapter my task is limited; I wish to examine what resources the advocates of a consequentialist moral outlook have to account for the felt demands of loyalty. In the end I hope to show the essential structure of consequentialism, particularly the conception of value characteristic of that theory, is inadequate for appreciating the legitimate and unique values expressed in rightful demands of loyalty. In order to do this I will examine some of the best cases that can be made from a consequentialist perspective in detail and attempt to show that even when a consequentialist claims to have found a consequentialist justification for a duty of loyalty, this justification fails to capture the essence of a demand of loyalty because consequentialism as a whole cannot properly acknowledge the value towards which that demand aims.

The Structure of Consequentialism

For my purposes here I identify as consequentialist any view that holds that the only relevant values for moral theorizing or deliberation are

embodied in states of affairs. These are among the class of views that Scanlon calls "teleological," and what distinguishes them "is not the elements of a state of affairs that they take to contribute to its intrinsic value (whether these include actions or only their consequences), but rather the idea that it is only states of affairs that have value" (Scanlon [1998], 80).1 What individuates a consequentialist view is its particular principle of the good—a theory describing an ideal (or at least preferable) state of affairs.

Consequentialism can be described in terms of two principles: a principle of the right (which governs what I ought to do, or how I ought to act) and a principle of the good (which determines what is of value). As I am understanding them, all classical consequentialist theories have more or less the same principle of the right (for moral agents): act in such a way as to bring about the optimal state of affairs as described by some given principle of the good (i.e., promote the good). Different species of consequentialism are distinguished by their different principles of the good—what they take to be the optimal state of affairs (Scanlon call this the "to be promoted"). For example, the principle of the good for classical utilitarianism (hedonism) is the state of affairs in which the net pleasure (total pleasure minus total pain) of the relevant subjects is maximized. Other forms of consequentialism have different principles of the good.

---

[1] It might be thought that by definition *consequentialism* would refer only to the subset of views that consider only the consequences of actions, and not the actions themselves. If this is taken to mean that what Scheffler calls "agent-centered restrictions" are excluded then this is generally correct; however, a view that gives some value to actions from an agent's perspective, as we will see in examining Scheffler's views, would still probably qualify as broadly consequentialist.

As we shall see when we examine Samuel Scheffler's attempt to retain a broadly consequentialist[2] picture of the good which is immune to an important objection against classical consequentialism, it may be thought possible to alter this principle of the right and still remain within a broadly consequentialist framework. However, no analogous alteration of the basic structure of the principles of the good is possible while remaining true to the essence of the consequentialist position. In other words, no strictly consequentialist moral theory can recognize as valuable anything not described in terms of a state of affairs. Thus, within such a view things have intrinsic value only, as Scanlon puts it, "as components of states of affairs—as things that *occur*, and that it is good (or bad) to have occur."

This is extremely important, because this claim will severely delimit the scope of what a consequentialist can recognize as valuable. Thus, for a consequentialist, any statement of the form "x is valuable" can only mean that the occurrence of x contributes more to the optimal state of affairs than the non-occurrence of x.

Consider, just as an example, Kant's assertion in the famous opening of the *Groundwork* that there "is no possibility of thinking anything at all in the world, or even out of it, which can be regarded as good without qualification, except a *good will*" (393). A consequentialist will think this statement is not only false, but also deeply incoherent. The consequentialist will think it false because he can point to occurrences that, according to his preferred theory of the good, he will assert to be intrinsically valuable. He will think it incoherent

[2] When I speak of Scheffler's project as "broadly consequentialist" with respect to the good, what I have in mind is that his account of value is still in terms of a Scanlonian "to be promoted." It may be that he does not wish to specify the exact nature of this to be promoted in terms of a metaphysical concept like a *state of affairs*. My discussion of Scheffler will be expanded below.

because a "good will" is not, and cannot be, one of those occurrences. This is simply because a "good will" is not an event; a will does not *occur* in the relevant sense.[3]

The mistake, according to the consequentialist, is in thinking that there is independent (intrinsic) value in having a particular orientation of the will. This is not to say, however, that all orientations of the will are equally valuable. Indeed, in the classical models a particular orientation of the will, as embodied in the principle of the right to promote the good, is the correct one to have. However, the *reason* that this orientation of the will is to be preferred has got nothing to do with it *qua* the will, but only because such an orientation will best promote the putatively optimal outcome.

What this means is that in certain cases it might well be that the best way to achieve a desired outcome is to cultivate orientations of the will that do not directly (from the agent's perspective) promote the good. Suppose, for example's sake, that our principle of the good describes the optimal state of affairs as one in which pain is minimized. Suppose as well that there is a brilliant but very selfish doctor who happens to be better able to ease the pain of others than anyone else but has no particular desire to do so. However, suppose he is motivated by a desire for money. In the short run, the best outcome—the minimization of pain—will not come about through my attempting to get him to embrace my principle of the right; to the extent that it is under my control, my best course of action would be to pay the doctor as

[3] Dick Miller has pointed out that there might be some kind of rather weird consequentialist who would take as principle of the good the state of affairs of wherein good wills (or souls) are maximized. This would be a very strange thing for a consequentialist to say for a variety of reasons, but most importantly because it is not clear that the consequentialist can give any kind of specification of what a good will is. The concept itself seems inimical to his implicit theory of value, and as such either incoherent, or, strictly speaking, arbitrary.

much as would increase his efficiency in relieving others' pain. In other words, the governing factor in what attitude it would be best for me to cultivate in the doctor is not the intrinsic value of the orientation of his will, but what will most effectively cause him to minimize aggregate pain.

Now a consequentialist might argue that in the long run I would better serve the cause of minimizing pain by cultivating in the doctor an attitude of respect for his patients and ultimately a direct appreciation of the good in minimizing their pain. However, there is no sufficient reason to think that this need be the case. It may turn out that by making the doctor less selfish I make him less efficient; he comes to identify and sympathize with his patients, directly desiring that their pain be minimized, but that causes him to spend more time with each patient, makes him less willing to make important snap judgments, and in general slows him down. In any case, in the short run, it becomes clear that given his psychological make up as it is available to me, the putative good can best be promoted by accommodating rather than changing his attitude.

The point of this example is merely that the classical consequentialist can attach no value to attitudes or orientations of the will except insofar as those attitudes promote or impede the occurrence of states of affairs. In fact, consequentialism considers states of the will only instrumentally and can argue for the promotion of any given state of the will only by reference to an efficiency criterion. Thus, as with our doctor, it may very well be the case that the optimific outcome is achieved only when some given agent not is a

consequentialist at all (i.e., he does not his actions according the principle of the right—viz., promote the good).[4]

The Demands of Loyalty in a Consequentialist System

Now that we have some picture of the structure of consequentialism we can return to the question at hand—can consequentialism account for the demands of loyalty as normatively legitimate? If the demands of loyalty are legitimate, then the values towards which those duties aim must be real values. Therefore, it is clear that for a form of consequentialism to account for the demands of loyalty as legitimate it must be able to recognize the values of loyalty as (independently) valuable. This means, given the formal structure of consequentialism as we have just examined it, that these values must be embodied within states of affairs. So we are left with a question that is relatively easy to state, if not to answer: can the values which inform the demands of loyalty be adequately understood as components of some state of affairs?

To answer this question will require us to have a reasonably clear understanding of the nature the values informing the demands of loyalty. One thing we might observe immediately is that it does not seem to be the case that the value of loyalty is properly understood as strictly additive. In other words, it is not simply the case that the more loyalty there is in the world the better. If this were the case, it would follow that I would have a duty to cultivate the

[4] This phenomenon has been explored by Peter Railton in his attempt to shield consequentialism from the charge that it is inevitably alienating ("Alienation, Consequentialism, and the Demands of Morality," 1984). I examine that discussion below.

maximum number of occasions for expressing loyalty. Further it would seem that I would have reason to be disloyal in certain occasions if by doing so I could maximize loyalty on the whole. 5

The point is that the value of loyalty cannot be properly appreciated if it is viewed strictly in terms of the occasions of behavior that express it. Furthermore, in the occasions where I do express loyalty I do not do so in order to maximize the occurrences of loyal behavior in the world, but because I care about the person or institution, etc. to which I am being loyal. Thus our question has now become more precise: can the value of the object of my loyalty be captured in terms of a state of affairs? Unless that value can be understood within it, consequentialism is systematically insensitive to real goods.

I want now to look at one attempt to construct these kinds of duties within a consequentialist framework, and then ask how well this attempt answers the questions we have posed. One thing that essentially characterizes the demands of loyalty is their particularity—these demands have both particular objects and limited (i.e., particular) scope. In the literature these kind of demands have come to be called "special duties" or duties of special concern; let us examine the attempt made by Robert Goodin and Philip Pettit6 to derive duties of special concern within a consequentialist framework.

---

[5] This is somewhat analogous to the case of friendship as examined by Scanlon: "We would not say that it showed how much a person valued friendship if he betrayed one friend in order to make several new ones, or in order to bring it about that other people had more friends" (89).

[6] "The Possibility of Special Duties" (1986).

Goodin and Pettit

Goodin and Pettit have attempted to develop a strain of consequentialist thinking that is in accord with common sense morality concerning what they call "relativised independent" obligations. In doing this they purport to examine and meet what they take to be the greatest obstacle to special duties, the "neutralist's challenge," by modifying their account of consequentialism to incorporate a "responsibility criterion." They claim that this serves to relativise duties to particular agents without fundamentally altering the character of duty. Examining this attempt will bring us some way towards answering the questions I posed above.

To begin we should note that they classify putative duties of special concern as "relativised independent" obligations. In other words, they are duties specific to the bearer that do not apply universally to everyone; they have an independent moral status and are not merely instrumental in satisfying some more basic universal duty.[7] Thus if there is a duty of special concern such as "Bob has a duty to educate his children," it is relative to him (i.e., Bob does not have this duty because everyone has a duty to educate Bob's children, but because they are *his* children), and independent in that it does not depend on any more basic duty (e.g., some kind of implicit promise

[7] In their usage, a given duty is relativised if it is of the form 'A has the duty to bring about some state of affairs specific to A,' e.g., Bob has a duty to educate Bob's children, while some other person Sam has a duty to educate, not Bob's children, but only Sams's. A given duty is dependent if it is instrumental to satisfying some more basic duty (e.g., Bob has a duty to abide by his contracts because contracts constitute a form of promising and everyone has a duty to keep his promises).

thought to have been made by Bob to his children).[8] With these distinctions in hand let us return to the problem.

Goodin and Pettit view the potential problem—the "neutralist's challenge"—to be the most pressing objection to putative duties of special concern. They think this because they claim that "special duties break the intuitive connection between the right and the good, for they make the duty of parents toward their child a function of something other than just the call of the good" (Goodin and Pettit [1986], 654). Of course, the force of this is most pressing when morality is viewed through a consequentialist lens, and those who fail to share the intuition, or more precisely this view of the good, are less likely to be moved to the same degree.[9] To see this let us look at the argument as Goodin and Pettit develop it:

> 1. It is (a) necessary and, given some general qualifying conditions, it is (b) sufficient for someone's having the independent duty that-*p*, that *p* should be desirable. Duty is, and is exclusively, a function of desirability.
>
> 2. If *p* is desirable from the standpoint of *A* – if *A* would be right to believe that is was desirable – then *p* is equally desirable from the standpoint of any assessor. Desirability is universal.
>
> 3. Therefore if *A* has the independent duty that-*p* then: by 1-a, *p* is desirable from *A*'s standpoint; by 2, it is desirable from everyone's standpoint; and by 1-b, it is something in respect of which everyone, given the appropriate qualifications, has a duty. (661)

[8] I am not claiming that this is necessarily a plausible more basic duty, merely that an independent duty does not depend on any kind of further duty.

[9] This is certainly not to say that deontological theories cannot generate a kind of neutralist challenge, but it will be somewhat different in character.

They note that there two ways of rejecting this argument—either, by rejecting (1) that duty is exclusive a function of desirability, or by rejecting (2) that desirability is universal. They claim that each of these responses is unsatisfactory.

They argue that the first strategy is most commonly "intuitionist in form" and "just uncritically endorses whatsoever ordinary intuitions we might have about our duties without offering any reason for supposing that these (and only these) are obligations which we should acknowledge" (663). Of the second strategy they say it "misrepresents the special duties it purports to save" (664). Their argument for this is that this view accepts "that if something attainable by an agent is attractive then it is potentially obligating and gives rise to a non-conclusive moral demand." But this would entail "that other people [could not] be in a position to find attractive the desideratum involved" in a given agent's special duty, since otherwise it would become a duty for them as well and so de-relativised. Thus, if Bob has a special duty to educate his children, this proposal would require that nobody else could be in a position to find the educating of Bob's children desirable, since it would then become a duty incumbent upon that person as well. However, it seems "absurd to suggest that other people" might not find a state of affairs such as Bob's children being educated to be desirable.

Their solution is really a modified form of the first response. They want to suggest that duty is not *exclusively* a function of desirability; rather it is (exclusively) "a function of desirability and responsibility, of attractiveness and accountability" (665). With the proper criterion for assigning responsibility they think they can avoid the neutralist's conclusion by relativising responsibility such that for some given state of affairs *p* (e.g., Bob's

children being educated), *p* can be recognized as universally desirable without this generating a corresponding universal duty for all who recognize the value of *p* to bring it about. This is because the responsibility to bring about *p*, via an application of the responsibility criterion, is Bob's alone (or Bob's and his wife's, etc.). Thus in the abstract anyone might find it attractive that Bob's children be educated, but the mere fact that someone acknowledges that this is a preferable state of affairs will not obligate him to bring it about, unless he happens to fall under the constraints of the responsibility criterion. Let us then look at Goodin and Pettit's proposed criterion.

> An agent *A* (be A an individual or group) is responsible for a state of affairs *p* if and only if:
>
> 1a. *p* is (virtually) uniquely susceptible to A's influence, whether that influence amounts to partial or total control;
>
> OR
>
> 1b. *p* is susceptible to the influence of *A* and a number of other agents; and it is not possible for those agents to exercise influence simultaneously without compromising the desired outcome *p* or some other desired result; and *A* is the salient one to assume control.
>
> OR
>
> 1c. *p* is susceptible to the influence of *A* and a number of other agents but it is possible for these agents to exercise simultaneous control without compromising *p* or any other desideratum;
>
> AND
>
> 2. A is in a position to know the truth of whichever one of those three conditions obtains. (666)

The purpose of this criterion is to assign responsibility most fundamentally to the person(s) in the best position to bring about the desired

state of affairs. This means that the good to be achieved (towards which the duty aims) is desirable in itself (making the duty independent) but binds only such persons as meet the requirements of the responsibility criterion (thus making it relativised). Thus, everyone can recognize that Bob's children being educated is a desirable state of affairs that generates corresponding duties, but only a small number of people (Bob, his wife, etc.) are directly responsible for making sure that this state of affairs is realized.

To see how this plays out let us apply this schema to a concrete example, the case of my special duty to see to the education of my children (the desired state of affairs in question is that of my children being educated, which we will name p). If there is such a duty then according to Goodin and Pettit a series of conditions will be true. First, it will be the case that it is desirable for my children to be educated. Second, an appreciation of the attractiveness of that state of affairs is open to anyone. Third, given that it is a special duty of mine, according to the responsibility condition I must either be: (a) (virtually) uniquely influential in bringing about p, or (b) the "salient one to assume control" among a group of agents any of whom might be influential in bringing about p, but who cannot all exercise that influence without compromising p or some other desired result, and aware which of these two cases obtains.

So which of (a) or (b) is likely to be true? If the desired state of affairs is my children being educated it seems strange to think that I alone could be influential in bringing this about. Since most children are not formally educated by their parents, it is reasonable to assume that there are a variety of people who might be influential in bringing it about that my children be educated. My wife most immediately comes to mind, as well as other

members of my family, immediate and extended. Family friends and neighbors also would exercise some degree of influence. Of course the authorities charged with public education could exert a great degree of influence. Presumably, therefore, the case in question would answer to condition (b) of the responsibility criterion; I am merely the "salient one to assume control."

This is the view that Goodin and Pettit express. In cases like providing "for the education of my children,… there is no realistic hope of smooth simultaneous control, so we search for the single most salient party,… [who] is likely to be me…" (670). They go on to say that under normal circumstances "information, resources and the actual conventions of familiar societies will dictate that I am the obvious candidate to be given control."

Under the description they have given, what is ultimately most important is just a desirable state of affairs. In this case, what is valuable—that from which the duty derives its import and meaning—is the state of affairs of my children being educated. The question we need to ask is whether or not the value of this state of affairs completely captures the value that underwrites the demand of special concern to favor my children over others in provisioning an education.

I want to suggest that it does not. Goodin and Pettit want to argue that the good towards which any duty aims must be, in principle, universally desirable. Remember they argue against making it a condition of duties of special concern that the good towards which they aim be unavailable to others on whom the duty is not incumbent. Another way of putting this is that a good is a good, whether or not any particular person recognizes this, and so if x is in fact a good, then x is a good for everyone in principle.

Now it is important to remember that they think "that if something attainable by an agent is attractive then it is potentially obligating and gives rise to a non-conclusive moral demand." They decline to allow the desirability of a good to be particular because they say it would be "absurd to suggest that other people will not find attractive what is at stake in the special duties that common sense ascribes to a particular agent." I think there is a confusion here to which Goodin and Pettit are insufficiently sensitive.

First I want to make a more general point about the nature of what constitutes a good—i.e., what is attractive.10 Suppose that I have two married friends, Anne and Bob. Suppose as well that I think either or both of the following: a) that marriage, in general, is a good, and so that, ceteris paribus, any instantiation of marriage is a good,11 and b) that Anne and Bob's marriage is a good. (Say, for instance, that I have known both Anne and Bob from before their marriage; I can see the love they bear for each other and how they each have greatly benefited from the marriage). In any case, I recognize that Anne and Bob's marriage is a good; I take it to be valuable; I take it to be attractive. However, it would be ludicrous to suggest that I value Anne and Bob's marriage the way (much less to the degree that) Anne and Bob each value their marriage and each other.

As I noted above, Goodin and Pettit suggest that it would be "absurd to suggest that other people will not find attractive what is at stake in the special duties that common sense ascribes to a particular agent." This is true—I, and other people, can find Anne and Bob's marriage valuable—but the point is

---

[10] Of course, to say that something is attractive does not mean what I just happen to fancy it; rather it means that I judge it to be a good.

[11] Though I do happen to think this, nothing hangs on it for the purposes of this example, and even should it fail to be the case I trust that the general point will nonetheless be apparent.

that we will not value it in the way in which Anne and Bob value it. As such, although it is valuable (both in principle and actually) to people besides Anne and Bob, it is certainly not the case that those of us who do value their marriage will each have the kind of duties of special concern that Anne and Bob bear to each other.

So we avoid the problem because we do not fall prey to the "absurd" notion that people besides Anne and Bob cannot find a real good (in this case, their marriage) attractive. Accordingly, although what I can do to promote their marriage is probably rather limited, it might seem that, since I am a friend of both Anne and Bob, I have some kind of *prima facie* obligation to do whatever I can. It might even be suggested that anyone who sees marriage in general as a good would be *prima facie* committed to promoting Anne and Bob's marriage to whatever extent he is able, even if he has no prior knowledge of Anne and Bob.[12] Thus, what I find attractive—i.e., the real good that is their marriage—can inform "a non-conclusive moral demand," but that does not imply that it informs the same moral demand that obliges Anne and Bob with respect to each other and their marriage.

---

[12] Of course, this is not to suggest that anyone who values marriage as a good has some kind of obligation to seek out all the married couples in the world and promote their marriages; rather, it means simply that to the extent that he happens to come into contact with Anne and Bob such a "non-conclusive" obligation might arise. So if I have a house to rent and based on all other factors I have narrowed down the potential tenants to two couples—one married and the other not—to the extent I esteem marriage a good I have *a reason* to favor the married couple over the unmarried one. In other words, though they are identical in other relevant respects, I make a non-arbitrary justified choice in favoring the married couple. It is only in situations like this, where some other more important factor may give rise to a conclusive moral demand in the other direction, that I have some kind of *prima facie but non-conclusive* demand to favor the value I esteem. I am not asserting that this is necessarily justified, I am only suggesting that valuing something like marriage as a good can generate demands to act in particular ways even towards people with whom one does not have a special relationship. Another example of a less specific *prima facie* obligation for him that values marriage in general might simply be to support a tax structure that doesn't "punish" marriage.

Thus the general point is this: a real good can be valuable in different ways to different agents, and thus inform different duties of special concern with regards to those agents, while remaining available, as a good, to everyone in principle. Goodin and Pettit wanted to avoid making attractiveness non-universal because to their minds a good, in order to be a good, must in principle be recognizably a good for everyone. What this example shows, however, is that even if it is true that a good is in principle recognizable as a good universally, it need not be equally *valued*, morally or otherwise, by everyone who recognizes it as a good.

This naturally suggests that if there is a demand of special concern on me to favor x, it is informed not merely by the fact that I recognize that the occurrence of x contributes to some good, but also by the fact that I value x in some particular way. What makes a duty of special concern *special* is not that the good towards which I am aiming is necessarily unavailable to others. Rather, what makes it special is that the value which underwrites the duty of special concern has a volitional component founded on the fact of my special relationship with the object of my special concern. As I suggested above when examining Kant's assertion about the good will in a consequentialist light, consequentialism, being wed to a view that the good can exist only in terms of states of affairs, is unable to adequately account for the volitional constituents of value.

To better appreciate this let us look again at some examples. I return again to Bernard Williams' case because I think it will shed more light on the nature of the value at stake. Recall that the situation involves a man being forced to choose between saving his wife or another woman from a fire, when he does not have time to save both. The question Williams is considering is

whether the fact that she is his wife is enough to justify the man's decision to favor her. He writes:

> … surely *this* is a justification on behalf of the rescuer, that the person he chose to rescue was his wife? It depends on how much weight is carried by 'justification': the consideration that it was his wife is certainly, for instance, an explanation which should silence comment. But something more ambitious than this is usually intended, essentially involving the idea that moral principle can legitimate the preference, yielding the conclusion that in situations of this kind it is at least all right (morally permissible) to save one's wife…. But this construction provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife. (Williams [1981], 18)

I fear that Williams' assertion "that his motivating thought, fully spelled out, would be the thought that it was his wife" may be somewhat misleading, and in any case some more spelling out—not of the thought, but of its significance—is in order.

Williams seems to think that the train of thought leading up to action in this case does not ever get to the stage of moral vetting. The "one thought too many" consists in the thought: "is what I'm about to do morally permissible?" Within the mental universe of the husband, seeing that his wife is trapped in the burning building, his thought should simply be the imperative: "save my wife." In fact, the question of "choosing" may never to come up. The train of thought is not this: 1) there are two women trapped in the burning building, 2) it is apparent that I can only successfully save one of them, 3) which one should I choose?, 4) Oh, I should save my wife. Even less likely would be included the "one thought too many": 5) But am I morally justified in saving

my wife? It seems more likely the thought is more on the order of: Oh no, my wife is trapped inside; SAVE HER NOW!

Now we have to be careful here. So far it might be conceded that this latter thought might best describe what a loving husband does think, but avoids the real question: is he justified in having this reaction; is he justified in not exposing his desire to save his wife to moral evaluation? In other words, does he not fail as a moral agent to the extent that he acts without first assuring himself of the moral permissibility of so acting? We are interested not only in the descriptive psychology of someone acting in this situation but also in the moral legitimacy of that action.

To make this objection, however, is precisely to miss the point. The question at hand is whether this kind of moral introspection is always the proper standpoint to have with respect to one's will. In the case at hand Williams seems to be suggesting that to subject the impulse to favor my wife to further moral evaluation—to have "one thought too many"—is precisely to fail to value my wife as I should value her.[13] To value her properly is to recognize that she is so central to my projects—to my character—and her good is so inextricably intertwined with mine, that my acting against her good is, to a large degree, volitionally impossible.[14]

---

[13] The claim here is not that when it comes to acting on my duties of special concern all moral evaluation is to cease; it is much more complicated than that. The point is that there are certain objects the value of which I can only properly express in having my will directed towards favoring them under most circumstances. This obviously does not mean that should I should favor my wife irrespective of any other consideration. Should she go on a wild murder spree I should not shield her from moral evaluation; not only do I have a duty to try to stop her, but also by the same token a special responsibility to do so.

[14] When I say that doing ~X is "volitionally impossible" I mean more or less the converse of what Harry Frankfurt means when he says that the doing of X is "volitionally necessary." These points, including an explicit discussion of Frankfurt, will be covered more completely in Part II, but the general idea is that an action is volitionally impossible if it utterly conflicts with my central self-defining commitments.

What is at stake here is the notion that "deep attachments to other persons," as Williams goes on to say, "will express themselves in the world in ways which cannot at the same time embody the impartial view." I want to suggest that they "express themselves" by directly conditioning the will of the agent. I will explore this in much greater depth in Part II, but at this point what we need to appreciate is that the human agency itself is founded at a very deep level on these kinds of attachments. Williams again, "unless such things [expressions of these deep attachments] exist, there will not be enough substance or conviction in a man's life to compel his allegiance to life itself. Life has to have substance if anything is to have sense, including adherence to the impartial system; but if it has substance, then it cannot grant supreme importance to the impartial system, and that system's hold on it will be, at the limit, insecure" (Ibid.).

What I want to take from this is the notion that the value which informs these demands of special concern is somehow intrinsically caught up in the volitional structure of agents in special relationships. It should be clear that what is efficacious in determining this value is not, for instance, that I have a marriage license. In other words, I am taking for granted at the moment, as I think Williams is as well, that the fact that the man is married to some woman is equivalent to his being deeply attached to her. Of course it is possible to be married to someone whom you do not love, but in this case a variety of points are in play. There may be a sense in which one is morally deficient in thinking of one's wife in that matter. In other words, one is failing to value something that is valuable. Additionally, however, if one's relationship is like this, then there might be little in the way of special concern at all, in which case there would be little in the way of a demand of special concern.

With this notion of volitional value in hand, incomplete though it is, I want now to go back to our original question for consequentialism which launched our investigation of Goodin and Pettit. Recall that that question was: can the value of the object of my loyalty be captured in terms of a state of affairs? We should now be able to see that the answer to this question is "no." The reason for this is simply that the value of the object of my loyalty in cases where the demands of special concern are conclusive is intrinsically bound up in my volitional structure—i.e., in my character. Furthermore, that value, the value intrinsic to my will, cannot be accounted for in terms of states of affairs; this value is not something that occurs in the world, but is bound up in the structure of my will itself.

Naturally, Goodin and Pettit are not going to be concerned about this directly because I suspect that they would simply deny that there is some unique morally significant value that is intrinsically intra-volitional. My response is that the phenomenology of the cases we are interested in (e.g., the loyalty exhibited by the man in Williams' example) cannot be properly understood except as cases of intra-volitional value. The problem for Goodin and Pettit is that they are wedded to a theory of value that is deeply inconsistent with our lived moral life.

Accordingly, *the kinds of duties that Goodin and Pettit generate are not, and cannot be, duties of loyalty at all*. It is absolutely vital that we see that the "special duties" that Goodin and Pettit purport to generate according to their "responsibility criterion" are not properly demands of loyalty, because the values that underwrite those duties are not specific intra-volitional values, but universal values made incumbent upon a given individual or set of individuals through an efficiency criterion. The responsibility criterion is

merely a device that under "normal circumstances, information, resources and the actual conventions of familiar societies" will most efficiently ensure the achievement of a good whose value is in no way uniquely available to him whom the criterion identifies as the "salient one to assume control" (Goodin and Pettit [1986], 670).

It is important to see that this failure is in no way peculiar to Goodin and Pettit. In fact their attempt is reasonably plausible as a way to derive a consequentialist duty that is superficially homologous to genuine demands of loyalty, but these consequentialist homologues cannot be genuine demands of loyalty precisely because they are not informed by a proper understanding of the value of the objects towards which genuine demands of loyalty are directed. The failure is not a lack of facility in their arguments but rather a misconstrual of the nature of the demands of special concern. This misconstrual is intrinsic to their consequentialist theory of value and its insistence that only states of affairs can have normative value. Instead, I think that the peculiar value of the objects of duties of loyalty lies significantly in the volitional structure of the agent and that this value cannot be properly accounted for in terms of a state of affairs.

Now this objection is most powerful against an act-consequentialist theory, whereby what an agent ought to do in any case is constrained by what would best promote the state of affairs described by his principle of the good. It is possible that a rule-consequentialist might allow for an agent, in particular cases, to be *motivated* by non-consequentialist considerations like love and loyalty. A rule-consequentialist (or even a "sophisticated" consequentialist of the type I will discuss below in connection with Railton) need not examine each particular action under a consequentialist calculus.

Thus, it might be argued such forms of consequentialism are not vulnerable to these sorts of criticisms arising from a Williams-type example.

However, I think this is wrong. Consider that a consequentialist must necessarily view these motivations only in terms of their utility—i.e., their tendency to promote some state of affairs. The consequentialist theory can never regard these considerations as themselves intrinsically valuable. They are only valuable insofar as, given the contingent conditions in which agents finds themselves, these motivations tend (for the most part and over the long run) to promote the desired state of affairs. If the relevant circumstances were to change in such a way that these motivations no longer had the effect of promoting that state of affairs, the consequentialist would be constrained to jettison them, as they have no independent value beyond their utility.

If I am right in maintaining the existence of intra-volitional values, then those values represent goods independent of their tendency (as informing the motives of agents having those values) to promote any given state of affairs. Those values are a function of the structure of the will, not of considerations extrinsic to it, such as some external state of affairs. I want to stress that a consequentialist can only recognize the value of love and loyalty extrinsically in terms of the effects they have when agents are motivated by them and not as valuable in and of themselves.

It would now be well to consider some issues raised above but not yet fully developed. My claim thus far has been the Goodin and Pettit's fundamental mistake is in no way unique to them, but is intrinsic to consequentialism. It is important to note that the criticism is founded on the claim that consequentialism simply cannot properly account for the values

that underwrite genuine demands of special concern. The claim is *not* that consequentialism cannot recognize these values at all, but that it cannot do so in the manner necessary for them to underwrite conclusive demands for action.

As I promised above, I want to look at Peter Railton's discussion of some closely related points. His discussion concerns the phenomenon alluded to above, that the consequentialist good might objectively be best served by not promoting a subjective identification with consequentialism. He has observed that this is analogous to what Sidgwick called the "paradox of hedonism" (the observation that those who make pleasure their direct goal are often much less likely to achieve it).

As I have already suggested, the key distinction that Railton makes is between subjective and objective consequentialism. It may often turn out that a consequentialist picture of the good can be objectively promoted most effectively by agents who themselves do not subjectively understand their actions as being underwritten by a consequentialist principle of the good. What is to be preferred, then, is a "sophisticated consequentialist" aware of this fact: "someone who has a standing commitment to leading an objectively consequentialist life, but who need not set special stock in any particular form of decision making and therefore does not necessarily seek to lead a subjectively consequentialist life…[one who] seems to believe he should act for the best but does not seem to feel it appropriate to bring a consequentialist calculus to bear on every act" (Railton [1984], 153).

Railton wants to shield consequentialism from the charge that it inevitably alienates agents from their deep commitments (e.g., special relationships) by forcing them to view all their actions through a

consequentialist lens. He wants to suggest that requiring agents to see all their actions this way would have the effect of *not* maximizing the good objectively. According to Railton, the sophisticated consequentialist will take an "indirect" approach,[15] and determine that, given the nature and psychology of human beings as they are (not merely as the consequentialist might like them to be), the objective good (the best state of affairs) can be best promoted by allowing individuals to focus on their own particular attachments. This makes possible an consequentialism that recognizes that certain "characters" and "cultivated dispositions" best conduce to the objective consequentialist good. These characters and dispositions can and should be developed in service of the ultimate consequentialist good.

For example, what Railton seems to have in mind is that having a disposition to value one's wife specially will be the kind of trait that goes into making a character of the type the prevalence of which among relevant agents maximizes the objective good (as embodied in a consequentialist principle of the good). Thus the good understood as objectively consequentialist is best promoted by individuals who subjectively value objects such as their close family and friends intrinsically and not only by reference to some larger, more abstract good.

Is this enough to answer my complaint that consequentialism cannot adequately express the volitional nature of value? I do not think so, precisely because this draws a sharp (and I think false) distinction between moral value and the sorts of values that undergird the demands of love and loyalty. Railton allows agents to be motivated by love and loyalty only to the extent

[15] This indirect approach, however, is not the same as rule-consequentialism.

that those motivations have an overall utility for promoting some state of affairs. Such motivations are not intrinsically the expression of a good will.

Under such a view there can be nothing intrinsically good about the will motivated by deep attachments. It is merely a contingent fact about human beings, that having such subjective motivations will best promote some putatively preferable state of affairs. For human beings differently situated (or for non-human agents, for example), such motivations might fail to conduce to the preferred state of affairs. In such a case Railton's sophisticated consequentialist would have no compunction about abandoning those motivations. There is nothing intrinsically valuable about them.

Railton's two-levels approach seems necessarily to draw a sharp line between directly promoting the good and being moved to act by one's particular deepest attachments. In particular, by demanding that value inheres only in states of affairs, advocates of this approach must deny that having a will of a certain shape has intrinsic value. Accordingly, the values represented by the demands of loyalty are necessarily foreign to this conception of value. On such a view, loves and loyalties can not be properly understood as valuable in themselves, but only instrumentally valuable for promoting some state of affairs independent of their occurrence. I think this ultimately does violence to some of our deepest intuitions that our loves are intrinsically valuable independent of their overall consequences.

Scheffler

At this point I wish to turn to another theorist who is making a similar move in the service of saving a broadly consequentialist principle of the good,

at least to the extent that it is understood as embodying what is "to be promoted." Samuel Scheffler, in *The Rejection of Consequentialism*, attempts to give some weight to a kind of first-personal valuing, while nonetheless retaining the legitimacy of acting to bring about the best outcome in any case. Carefully examining this attempt will help us further clarify the central issue—viz., the nature of the values informing the demands of special concern.

In *The Rejection of Consequentialism* Scheffler examines two main objections to consequentialism. The first objection, raised by Bernard Williams, Scheffler calls the objection from "personal integrity." The second, perhaps as old as consequentialism itself and strongly reiterated by Rawls, Scheffler calls the objection from "distributive justice." Although it is certainly of general interest, the second objection and his response will not greatly concern us here. My focus will be almost entirely on the objection from personal integrity and Scheffler's attempt to meet it, for its applicability to the examination of duties of special concern is readily apparent.

Let us begin by examining the nature of Williams' objection, which we will find is of a piece with the complaint about alienation that Railton considers, and then consider Scheffler's response to it. Williams' objection arises because the classical consequentialist principle of the right requires agents to act in such a way as to secure the optimal state of affairs as described by some principle of the good. What this means at the most fundamental level is that agents are responsible for bringing about a state of affairs, and not merely, for example, for the orientation of their own wills. Since I am required to bring about a state of affairs it is easily conceivable that situations may (and in fact almost certainly will) arise in which I am forced to give up my own particular plans and projects because circumstances are such that by so doing I

am better able to bring about the preferred state of affairs according to the operative principle of the good.

Scheffler quotes Williams from Utilitarianism For and Against: "how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because someone else's projects have so structured the causal scene that that is how the utilitarian sum comes out?" This would be "to alienate him in a real sense from his actions and the source of his action in his own convictions" (Smart and Williams [1973], 116). The basic idea here is that since, according to utilitarianism (though this would presumably apply to any non-egoist consequentialism), the agent is responsible for promoting a state of affairs, circumstances will often require him to give up the projects that are central to his life (at least in his own self-construal) and force him to act in ways he would not otherwise choose, simply because the world is set up in such a way that doing this will better conduce to the optimal state of affairs.

Thus the objection is that utilitarianism robs the agent of personal integrity by requiring him always to be ready to sacrifice his own personal projects and plans at any time when doing something else would better conduce to the optimal state of affairs. It doesn't matter how important his projects are to him if giving them up will produce the better state of affairs. In other words, his ownership of his own life is threatened, forcing him always to respond to outside circumstances by taking a third-personal view of the situation and according his own projects no more consideration than anybody else's. Because at any time he can be required to drop whatever he is doing, no matter how important to him, to do something else that better promotes the desired state of affairs, he cannot pursue a life of his own—he cannot make

and execute the kind of long-term, life-shaping plans that make a life unique and worthwhile.

Scheffler considers this objection to be telling against classical consequentialism. He recognizes the importance of what he calls "the independence and distinctness of the personal point of view" and that classical consequentialism threatens it. Therefore, he proposes what he calls a "hybrid" consequentialist theory which incorporates an "agent-centered prerogative," but which nonetheless does not go so far as to demand the kind of "agent-centered restrictions" that mark deontological theories.

As a lengthy discussion of Scheffler's analysis of deontological theories will not be profitable here, let me merely note that Scheffler believes agent-centered restrictions—"restrictions on action which have the effect of denying that there is any non-agent-relative principle for ranking overall states of affairs from best to worst such that it is always permissible to produce the best available state of affairs so characterized" (Scheffler [1982], 2)—are not warranted by the independence of the personal point of view. On the other hand, the personal point of view can and does, on Scheffler's analysis, underwrite "agent-centered prerogatives" the inclusion of which makes his hybrid theory unique.

Scheffler believes that agent-centered restrictions (e.g., it is always wrong to kill an innocent person, even to save the lives of a greater number of other innocent people) are irrational. I do not want to rehearse his argument, but merely note this fact by way of explanation for why he favors agent-centered prerogatives instead. In essence, Scheffler wants to leave open the possibility that agents always may, but are not required to, act in such a way as to bring about the best state of affairs as characterized by some principle of

the good. This is distinguished from classical consequentialism which requires agents always so to act. Furthermore, this is distinguished from deontological theories incorporating agent-centered restrictions, which will hold that that there are some occasions when an agent is required not to act to maximize the consequentialist good (e.g., as when bringing about that state of affairs would involve an assault against an innocent person).

Thus, Scheffler thinks that a tenable moral theory must allow for the possibility of acting to bring about the best state of affairs as described by a given principle of the good in all cases, and yet not require the agent to do so. In certain circumstances the agent could choose not to act to bring about the best state of affairs in order to pursue his own plans and projects. This shields the hybrid theory from Williams' objection from personal integrity. Nevertheless, the agent could choose to do precisely what classical consequentialism requires, either because that just is his central personal commitment, or because he is willing, in this case, to give up his own personal projects for the greater good (a kind of consequentialist martyrdom). In any case, however, the optimal state of affairs is in principle knowable—there is always a non-agent-centered perspective from which it is possible to evaulate potential outcomes vis-à-vis some given principle of the good.

What makes the hybrid theory unique is Scheffler's inclusion of the "agent-centered prerogative."[16] As Scheffler understands it, the agent-centered prerogative should meet two requirements:

[16] His complete theory also address the second objection mentioned above by being "distribution sensitive," but this fact is less important for our purposes here and so will not be discussed in what follows.

> First, such a prerogative should not merely permit an agent to devote energy and attention to his projects out of proportion to the weight from an impersonal standpoint of his doing so, but rather it should do this in such a way as to permit the coherent integration of the agent's values and actions within the structure of a unified personality. But second, an acceptable agent-centered prerogative should place appropriate restrictions on the values and actions whose coherent integration and development it will protect (19).

Thus the agent-centered prerogative is sensitive to the first-person standpoint, by limiting the agent's responsibility to bring about the best state of affairs as viewed from an impersonal standpoint. Accordingly, agents are allowed "to devote energy to their projects out of proportion to the weight sub specie aeternitatis of their doing so" (22).

This agent-centered prerogative is important because it actually serves to specify, though still to an insufficient degree, the extent to which Railton's "non-moral" values can come into play in determining right action. The prerogative, were it fully spelled out, might give a principled distinction between those personal values which are shielded from being over-ruled by consequentialist ones and those that must be surrendered.

By allowing the agent to choose, "within certain limits," to ignore the demands of the principle of the good by favoring his own projects out of proportion to the weight they would otherwise receive, the hybrid theory allows the agent to preserve his own integrity by sometimes refusing to act so as to produce the optimal outcome according to his principle of the good. Now obviously, the proviso within certain limits is of central importance. Unfortunately, Scheffler's discussion of this key point is incomplete at best.

The problem is readily apparent when we consider what limits might be set on the prerogative. It would not be enough to shield all and only those

projects which the agent construes as central to his self-conception. This would not be narrow enough because the agent who has a particularly egoistic self-conception might claim an agent-centered prerogative to more or less ignore all outside moral demands simply by making such egoism central to his own life's project, and thus making a violation of it a violation of his personal integrity.[17]

Furthermore, to the extent that only those projects taken by the agent himself to be important are shielded, the theory would fail to recognize as valuable any value that a particular agent failed to embrace. The fact that some given agent does not properly value something (e.g., a man who abuses his wife) certainly cannot be taken to mean that the value in question is not intrinsically valuable.

Thus it seems as if the limits cannot be self-set, nor merely responsive to particular self-conceptions. Rather, it seems as if the legitimate limits to the prerogative could only be set by some reference to what projects are valuable ones to have, limiting protection to those projects, and excluding others that are too egoistic, narcissistic, or selfish.

At this point we can begin to see the problem with Scheffler's hybrid view. In order to meet the objection from personal integrity the hybrid view incorporates the agent-centered prerogative including a proportionality principle which allows the agent to give more weight to his own (legitimate) projects than he would be able to under classical consequentialism. However, this proportionality principle itself is under-described in Scheffler's discussion, and, as I will now try to show, this is no accident.

---

[17] In such a case the violation of his personal integrity would not only be non-objectionable from the moral standpoint, but would be in some sense required.

The problem, in a nutshell, is that the proportionality principle must itself constitute an independent principle of the good that is hostile to the conventional consequentialist principles of the good. Thus, I claim that any attempt to incorporate a agent-centered prerogative of the type Scheffler describes into a broadly consequentialist theory is bound to fail, because the differential responsibility level the prerogative is designed to secure is intrinsically incompatible with consequentialism.

Let us now try to see why this is so. The essence of any species of consequentialism is the particular principle of the good that underlies it. Indeed, Scheffler's own commitment to a broadly consequentialist theory of the good seems to reside precisely in the intuition that there must be some best state of affairs, and accordingly that we are duty-bound to bring it about so far as we are able. Thus a particular advantage of consequentialism is that it envisions the good universally—the good will not be unique to any given individual or group of individuals, but will consist on one, agreed upon, objective state of affairs in the world.[18] This is, of course, what was motivating Goodin and Pettit to assert the universal desirability of genuine goods.

Now if Scheffler's agent-centered prerogative were integrated into a consequentialist principle of the good, the resulting hybrid consequentialist theory would lose a univocal, objective conception of the best state of affairs, its main advantage. Consider two agents, A and B, who are in morally identical choice situations and faced with the choice to ϕ or not to ϕ. Let us

[18] This only holds in principle, of course. This would be the case under conditions of perfect knowledge of both the world as it is, and the results of any actions within it (much like Adam Smith's perfect market depends on perfect knowledge and zero transaction costs). In the real world in which neither of the conditions obtains, consequentialism must operate on best guesses about both how the world is and the likely consequences of any particular actions. Thus, in principle consequentialism is beholden to the social sciences—and this by itself must constitute a huge objection to it as a *moral* theory, but that is an argument for another day.

specify that a classical consequentialism, according to the principle of the good they both share, dictates that an agent in A's (and B's) situation is required not to ϕ. Now it is the case that A has a set of central commitments and convictions such that he is greatly invested in ϕ-ing, and such that not ϕ-ing would alienate him from those projects. However, B, while he might prefer to ϕ as it would result in some minor pleasure to him, is not invested in it to the degree that A is; so while he would be a bit put out by missing the opportunity to ϕ, B would nevertheless not experience anything like the assault on personal integrity that A would experience in giving up ϕ-ing. In this case, it would seem as if A should exercise his agent-centered prerogative to ϕ anyway, while B should not. Thus, despite being in morally identical situations, the optimal state of affairs for A and B would be different. Thus, the univocal nature of the consequentialist theory is broken.[19]

The point is not merely that differing preference schedules might change the consequentialist calculus for different agents; any act-consequentialist theory would take those kinds of circumstances into account. Rather the point is that Scheffler's theory breaks with the classically consequentialist commitment to the universality of the good, as the grounds for duty. Goodin and Pettit were clear about this when they sought to defend the universal desirability of the object of any duty (recall that they argued that for any action to be a duty, the good at which it aims must be universally desirable, even if responsibility for bringing it about is not universal).

[19] If it is not obvious, another restriction applies—viz. that the choice of whether to ϕ lie within the aforementioned, vague "certain limits." Thus ϕ-ing probably is going to be something like supporting the local opera company instead of sending the funds for famine relief, where A is a true devotee of opera and B mildly gets a charge from the end of Act I of *Turandot*.

Now I take this to be a telling objection to Scheffler's hybrid theory precisely because in attempting to save a broadly consequentialist theory of the good, he is pre-supposing that there will always be a unique best state of affairs (as described by some given principle of the good). In the situation described, however, we now have two different "best" states of affairs, dependent only on the pre-existing commitment of one of the agents to his particular projects. At this point, then, we have to wonder if the hybrid theory really is a kind of consequentialism at all.

This point can be expanded when we consider what it means for there to be multiple "best" states of affairs. If the hybrid theory is consequentialist then Scheffler is committed to arguing that there is a best-state-of-affairs-for-A and a best-state-of-affairs-for-B. Since consequentialist theories are individuated by their principles of the good, and each principle of the good describes a unique best state of affairs, the fact that we are presented with two different "best states of affairs," seems to imply that A and B have different principles of the good (because a consequentialist principle of the good simply describes his preferred state of affairs). In other words, the very fact of the existence of divergent "best" states of affairs for A and B, despite being, by postulation, in morally identical choice situations, implies that they must be operating with different principles of the good.

For different agents to desire to promote different putatively optimal states of affairs is for those different agents to no longer share in the universal desirability of their individual goods. As such, the different agents simply have separate, mutually incompatible principles of the good. Accordingly, they cannot motivate conclusive reasons against each other for claiming that certain particular actions are required by duty, *simpliciter*. They are simply no

longer operating within the same constraints and neither can offer sufficient independent reasons to conclusively compel the other to accept his preferred state of affairs. Accordingly, it does not seem as if either can conclusively persuade the other that he has a *duty* to perform some action that is necessary to promote the first's desired state of affairs, because for a consequentialist, duty is grounded on universal desirability.

If they end up operating with different principles of the good, yet they both initially subscribed to the same principle of the good (by stipulation), this implies that whatever difference there is between the two agents itself has moral content and so constitutes a kind of principle of the good in its own right (since it serves to modify their initially shared principle of the good). As we know that the difference was produced by the proportionality principle that allows A to place more weight on ϕ-ing than B, I take this to mean that this proportionality principle must then incorporate moral content, independent of the moral content contained in the original principle of the good. This is simply a complicated way of saying that the proportionality principle is itself a kind of principle of the good, whose distinctive moral content is just that making moral choices in accordance with one's own commitments and projects is *morally* valuable.

Well now we have two separate principles of the good, and indeed we will come to have as many as there are different individuals with different worthwhile projects protected by the agent-centered prerogative. Ultimately, then, the hybrid view begins to loose whatever advantages it originally had in claiming to be a kind of consequentialist theory at all. Furthermore, this still leaves completely unexplored the additional difficulty of specifying what projects will count a morally worthwhile.

Maximizing Good-willing as a State of Affairs?

I want now to clarify some of my claims regarding the fundamental incompatibility of consequentialism with the sort of volitionally constitutive valuing that is at the heart of my claims for loyalty. In particular, I want to be clear about why consequentialism simply cannot subsume my concerns in a more expansive understanding of in what the good state of affairs consists. It is important for my argument to see that the volitional values at the heart of loyalty are deeply incommensurate with the metaethics of value that underpins the consequentialist position.

A reasonably natural objection to my position, and one which I think both Railton and Scheffler would probably be inclined to make, is simply: why could consequentialism not subsume any sorts of volitional values within a principle of the good, simply by describing the desired state of affairs as being partially constituted by good-willing? In other words, why cannot the desirable state of affairs simply be that state in which good-willing occurs (or, even, is maximized)? Would not such a principle of the good obviate my concerns by making room for good-willing, and yet still preserve the essential metaethical conception of value intrinsic to consequentialism?

To answer this requires us to think about how the value of the good-willing is to be understood. In a case like the one proposed, why exactly is any instantiation of good-willing (whatever it may be) good? For simplicity's sake let us keep to our staple example of the man loving his wife, and ask,

assuming that it is good,[20] what makes an action expressing a man's love for his wife good? Within consequentialist metaethics there can be only one answer: such an action is good just insofar as it promotes the desired state of affairs. The point holds generally, within a consequentialist metaethics, for any given consequentialist theory, the moral goodness of all goods is derivative from the state of affairs prescribed by that theory's principle of the good.

This means simply that no instance of good-willing can be intrinsically good in-and-of-itself. Rather its goodness is entirely derivative. It my be objected, however, that this ignores the fact that we specified that the state of affairs to be promoted was in fact partially constituted by the occurrence of good-willing. Therefore, while its goodness is derivative, it is derivative from that which it itself partially constitutes and to that degree is valued intrinsically. In other words, while it may not be valued intrinsically in-and-of-itself, it is still intrinsically valuable in that occurrences of it contribute to the state of affairs from which all goods derive their value.

It is very important at this point to see that what is counted as good is the occurrence of an instance of valuing. This is simply because the components of a state of affairs can only be occurrences, events in the world. As such, we need to be clear that the putatively good state of affairs in our sample consequentialism is constituted not by good-willing itself, but by occurrences of good-willing. I suspect many consequentialists would be apt to reply that this is a distinction without a difference, but I think short consideration will prove otherwise.

---

[20] For the purposes of our discussion I am taking it for granted that it is a case of good-willing to love one's wife; however, as will become apparent nothing turns on this. The reader may substitute whatever example of willing as is pleasing.

The difference between good-willing and the occurrence of good-willing is the difference between doing an action and the effects of that action (i.e., its consequences). The distinction essentially involves the difference between a first-personal and third-personal conception of value. The value expressed in the good-willing is intrinsically bound up in the act of willing, not with the consequences of that act. Thus, a volitional value understood third-personally is divorced from the agency of the agent.

Perhaps this will be clearer if we consider how a non-consequentialist metaethics might construe the goodness of an instance of good-willing. On one such account, what makes my good-willing good has to do intrinsically with the orientation and constitution of my will and nothing to do with the consequences it has. The value of the good-willing is internal to it as a willing, as expression of a particular orientation of the will, and does not derive its value from the state of affairs that eventuates from that act.

There are two clear, but very different examples of this sort of volitional value. On a Kantian understanding, what makes a will good is that it be determined only by a respect for the Moral Law, which means that is determined only according to a rational law which it gives to itself (and all other rational creatures in the Kingdom of Ends).[21] What makes a good-willing good for Kant is that the will have the proper shape, namely that it be internally constituted by and through a kind of rational respect, the form of which his entire moral philosophy is dedicated to describing. What I want to take from this, at the moment, is merely the idea that the goodness of a good-

[21] I will be discussing Kant's views at length in the next chapter, so I will not delve into detail here. I only wish to draw a contrast with the consequentialist metaethics of value, so the details are not vital to seeing the point.

will does not derive from the consequences of its actions in any respect, but is a formal property of that will.

The second example is similar in that the goodness of a good-willing is prescribed by a formal property of the will. This is the classical Christian (though not exclusively Christian) notion that a will is good only insofar as it reflects (is in concord with) the will of God. Again, at this point I am not interested in whether this is true or not. What I want us to see is merely that the goodness of a good will on such a view is a formal property of the will, a property of its shape or constitution. The effects and consequences of the act of will do not determine its goodness.[22]

There is deep disagreement here about the metaethics of value. At some point the disagreement will reach bedrock intuitions about what the good consists in. What I hope to argue here is that consequentialism simply cannot contain the intuition that what makes a good will good is a formal property of that will, of its shape or constitution. The reason for this is that the only value consequentialism of any sort can embrace lies in the occurrence or non-occurrence of some effect, not in the willing itself.

This may become clearer if we look again at Railton's "sophisticated consequentialist." Recall, a sophisticated consequentialist is objectively committed to promoting some consequentialist principle of the good while nonetheless allowing that he need not subject his every act to a consequentialist calculus. Thus, such an agent might be committed to certain values (such as loving one's wife intrinsically) that in any given case do not

[22] This is not to fall into the trap of saying such consequences are irrelevant. Surely the consequences of an action are important in understanding the relationship of the agency to the world, and as such can certainly inform the deliberations, etc. that often go into an agent's determination of his will. However, the point remains that the goodness or badness of the willing itself is not derived from the state of affairs that eventuates from it.

maximally promote his desired state of affairs. The reason for this is that, in the long run, the desired state of affairs overall is best promoted by allowing agents this lee-way. In other words, it just turns out to be the case that allowing agents to shield certain decisions from the consequentialist calculus will best conduce, on the whole, for all of society, to the desired state of affairs.

I think it is important that we recognize that these volitional properties are only valued contingently here. It just turns out that allowing agents this subjective non-consequentialism conduces overall to the best outcome, but there is nothing about the act of will itself that is valuable on such a view. If for whatever reason it turned out to be the case that having a will organized in this way did not conduce to the best outcome overall, then it could (indeed should) be abandoned without remorse, for nothing intrinsic would be lost.

For example, suppose it turns out to be the case that in our present circumstances the state of affairs of having children be educated is best effected by their being raised in traditional families with both parents. If this is the state of affairs we wish to promote, then agents would be allowed, under Railton's theory, to have attitudes characterizing such traditional families because overall this would best conduce to the desired state of affairs. However, suppose that the material conditions change such that, because of increased wealth or technology, etc., it came to pass that the state of affairs desired could best be achieved through compulsory boarding schools which separated children and parents from an early age. In that case, nothing internal to the theory would hinder this sort of social reorganization.

Nothing turns on the truth or falsity of these claims about what would best conduce to the desired state of affairs. Rather, the point is that by conceiving of value entirely in the occurrences of events, the particular first-

personal value of an act of volition cannot be adequately comprehended. Because the act of will itself is not a source of value in this metaethics, when contingent circumstances change the likely outcomes and effects of certain sorts of volitions, then the value of those volitions itself is changed and even perhaps destroyed.

I fear the relationship of these points may be getting a little bit obscure at this point. The objections I have leveled so far have got nothing in particular to do with loyalties, and can be shared by any sort of deontological moral theories, be they Kantian or Thomist, or of some other sort altogether. For the point so far is merely that the moral value of willing itself cannot be adequately captured by the consequentialist metaethics of value.

As will become more apparent in the next chapter on Kant and in Part II, I believe that loyalties essentially embrace a volitional metaethics of value. In other words, I argue that the value of loyalty consists in having a certain sort or shape of will—a volitional orientation of a certain kind. The defect of consequentialism lies precisely in the fact that it cannot see this value for what it is, but must "translate" it into terms of occurrences which conduce to certain states of affairs. It is only the states of affairs that have value, and the value of any particular volitions lies only in their effects.

Describing a state of affairs as constituted by good-willing does not get out of this problem, because the constituents of states of affairs are occurrences—the effects of willing in the world. It is not the act of will itself that is valuable, but merely its occurrence and the effects that follow from it. The good is thus never seen from "inside" the agent, but only from "outside"—from the way the world is or is not.

In later chapters I will have more to say about these topics, but for present purposes I want it to be clear what I am claiming for loyalty. I have not yet even tried to make a positive argument for the legitimacy of loyalties. Rather I have concentrated on trying to show the deep incompatibility of the consequentialist metaethics of value with the sort of volitional value which is at the heart of the proper understanding of loyalty. Thus, as of yet, nothing I have said here should be construed as my full argument for legitimate loyalties, only an argument against the possibility of consequentialism as a complete theory of normative ethics.

Conclusion

Neither Railton or Scheffler tackles commonsense morality head on, which not only asserts that in Williams' case I am "justified" in saving my wife, but that to fail to save, or even for that matter, even to delay to consider whether I am justified in saving her, is to act wrongly. The proper valuing of a wife by her husband that ought to motivate action in this case simply cannot be reconciled with a theory that dictates that all possible moral values must be embodied within a state of affairs.

Scheffler makes a game attempt through his agent-centered prerogative to offer a principled way to create a protected zone within which agents' life-constituting personal projects are shielded from the demands of impersonal morality, but even if that prerogative could be fully spelled out, it could do nothing more than beg important questions. The prerogative would have to adjudicate between conflicting demands, both the demands generated by

impersonal morality and the demands generated by the agent's own central life commitments.

In order to do this, however, the prerogative would seem to require some kind of principle that identifies which of the agent's life commitments are legitimately to be protected in which cases. It seems strange to think that a *non-moral* principle could allow us to make such determinations. However, if it is a moral principle, then how is that to be reconciled with the over-arching broadly consequentialist principle of the good that is supposed to underwrite the demands of impersonal morality? In other words, if the consequentialist principle of the good does not have universal scope, then it is unclear how it can actually *be* a principle of the good.

To modify the principle of the right in the way Scheffler tries to do with his agent-centered prerogative inevitably introduces moral content into the practice of practical reason that is independent of the consequentialist principle of the good—viz. that it is valuable to express my own central commitments. But why is this valuable? This cannot be expressed adequately in terms of states of affairs, because the value in question is one of my orientation with respect to my own will; that it be truly *mine*.

It is these volitional values that no consequentialist theory, "sophisticated" or "hybrid," Railton or Scheffler, can adequately comprehend, precisely because the values are simply alien to states of affairs, or the "to be promoted." I cannot "promote" a volitional value by any other means than by willing it. And though having such a conditioned will naturally entails that I am more likely to do some actions than others, the value of those actions does

not consist entirely in their *occurrence*, but lies partially in the fact that *I* did them, that they are the external manifestations of my immanent will.[23]

Accordingly, to the extent that the demands of special concern are intrinsically and inescapably connected with values expressed volitionally, consequentialism is inadequate as a theory of value concerning them. Of course, it remains true that this is the case only insofar as these values of special concern are real values, and for the theorist inclined to dismiss them altogether, this argument will fail to be compelling. Although defense of these values as real values must await Part II, I think we can see already that merely dismissing them is inadequate; rather, the theorist must try to explain away their appeal. However, I have tried to show in this chapter why such attempts are themselves uncompelling, both internally, and as descriptions of the relevant common sense moral phenomena.

[23] One thinks here of Hegel's thought that the will is "externalized" through shaping the world to its demands. The value of this, for Hegel, clearly lies not only in that the world is better ordered, but that the will (or self) in question develops and indeed is brought into being through this process.

CHAPTER IV

Kant and "Kantian" Approaches

In the last chapter I explored the resources available to adherents of broadly consequentialist moral theories to account for the demands of love and loyalty. My intention there was to show that they fail to adequately capture the volitional value of the proper objects of love and loyalty because of the inherent limitations of the consequentialist understanding of value. In this chapter I will turn my attention to the prospects for broadly "Kantian" approaches, the result of which will have a similarly negative conclusion.

I use the scare quotes around 'Kantian' advisedly; while I will consider Kant's view directly, I also want to cast my net wider to consider views that have been called, justifiably or not, "Kantian." Thus, I am interested not only in applying Kant's own thinking to the problem, but also in considering "contractualist" views of morality that, while not Kant's, nonetheless grow out a certain liberal tradition one of whose central foundations is Kant. As with the consequentialist views we considered in the last chapter, I believe it will turn out that all these Kantian views will be inadequate (though one may be superior to the other), but in an interestingly different way.

Kant

Let us begin with Kant's actual view, and once that is in place it will be easier to appreciate the moves that characterize some of the "Kantian" alternatives. I want first to discover how the demands of love and loyalty would come out in Kant's moral theory as expressed most directly in the *Groundwork of the Metaphysics of Morals*. To do this we will have to have lay some groundwork of our own.

It is difficult to speak of "Kant's view of morality," for there are many live exegetical disagreements. There is a line of Kant scholarship that holds that Kant's thinking on the metaphysical grounds of morality changed significantly between the *Groundwork* and the *Critique of Practical Reason*. On this view it is thought that Kant himself came to realize the hopelessness (or at least superfluity) of the project of the Third Part of the *Groundwork*, in which he tries to derive the Categorical Imperative from pure practical reason alone (as "an a priori synthetic practical proposition") in a manner analogous to his derivation of the categories of the understanding in the *Critique of Pure Reason*. Recognizing this, so this interpretation goes, Kant pulls back and instead concentrates on the truly practical (i.e., directly action-guiding) aspect of his Categorical Imperative, without worrying overmuch about its supposed grounding in the nature of reason itself.[1]

---

[1] This kind of reading of Kant (with many variations, of course) has been especially popularized by Rawls and those following his lead. Consider this random example: on reading Rawls' "Themes in Kant's Moral Philosophy" (Förster [1989], 81-113) one might almost be led to wonder whether Rawls's copy of the *Groundwork* was missing Part III, as he makes no mention of it or the important themes therein. One might also wonder how far some of these interpretations (e.g., Christine Korsgaard's work) seek to "help" Kant out by saving him from himself, so to speak. In other words, many readers of Kant have been more impressed by (e.g., the universality of) the Categorical Imperative than its putative derivation from pure practical reason itself, and so have been inclined to stress the former and de-

An alternative view, which seems to me the more accurate as a reading of Kant, is that there is a fundamental consistency between the methodology of the critique of pure practical reason in *Groundwork* III and Kant's later works of moral philosophy. Further, any apparent inconsistencies arise not so much because Kant's thinking about the metaphysical foundations of morality changed, but merely because the focus of his discussions is different. On this view, it is not so much that Kant changed his mind, or gave up on the earlier project, but that he turned his attention to more practical matters in ethics, thinking he had already secured the metaphysical foundations in the *Groundwork*.[2]

Not being a Kant scholar I do not want to wade directly into the middle of this debate. However, I do think that it is impossible to come to a complete view of the majesty of Kant's moral project without taking *Groundwork* III seriously. For even if that project is ultimately a failure, it nonetheless represents the most sophisticated sustained attempt to derive the ultimate principle of (practical)[3] rationality—the Categorical Imperative—from the nature of reason itself. Further, any view that fails to come to grips with Part III simply cannot be fully accurate to either the spirit or the letter of the Kantian project.

---

emphasize, if not downright disown, the latter. Personally, I have serious reservations about such moves, which I think not only inaccurate to Kant's fundamental project, but also much less interesting. Such interpreters gain some additional *prima facie* plausibility, but at the expense of depriving the Kantian view of not only its most distinctive elements, but also its only real hope (in my view) of being correct. The fact that (again in my view) the full Kantian project ultimately fails suggest that these modified "Kantians" have even less hope for success.

[2] The leading defender of something like this view is probably Onora O'Neill, *Constructions of Reason*, especially Chapter 3, "Reason and Autonomy in *Groundwork* III" (O'Neill [1989]).

[3] O'Neill would go so far as to say, "that the Categorical Imperative is the supreme principle of human reason" (O'Neill [1989], 52), *tout court*, as it were.

Let us begin with a brief outline of Kant's project. Fundamentally, Kant's moral theory is an attempt to articulate how the requirements of the intelligibility of moral action condition which of such actions are morally permissible. Kant asserts that an inescapable requirement of practical rationality is the ability to give reasons for one's actions. This means that for my activity to be *motivated*—to exist within the sphere of moral reasons—I must be able to give maxims describing my actions in terms of means-ends reasoning. Thus, two questions can arise: 1) are my chosen means a rational way to obtain my given end, and 2) is my end itself a good end? These two questions correspond roughly to the two types of imperatives "hypothetical" and "categorical."

Hypothetical imperatives are conditional imperatives. They command certain actions as the necessary means for the achievement of a given end. A hypothetical imperative is in some sense neutral as to ends; it commands what it commands in virtue of a *given* end, whatever it may happen to be. The Categorical Imperative on the other hand commands what it commands as an end in itself. The Categorical Imperative is thus not conditional upon any further end.

Kant sees the space of moral reasons as informed by "maxims." It is by placing our actions in terms of maxims that describe our means-ends reasoning that we make those actions morally intelligible. This is because the maxims are law-like in form, specifying the relation between agents as free causes, and their actions as effects. In a footnote at 401, he defines that a "maxim is the subjective principle of volition; the objective principle (i.e., that which would also serve subjectively as the practical principle for all rational

beings if reason had complete control over the faculty of desire) is the practical *law*."

What is behind this is a need to find law-like relations in human moral activity. In other words, Kant thinks that a necessary requirement for the intelligibility of the moral realm is that it be governed by laws in just the same way that the intelligibility of the physical realm requires that it be governed by the laws of nature. He thinks that we are able to understand the physical world—that is, to ascribe regularity and predictability to it—precisely because it is not random, but governed by exceptionless laws. It is by discovering those laws that we are able to gain knowledge about the physical world.[4]

In just the same way, Kant thinks that the intelligibility of the moral universe depends on its being governed by laws. Accordingly, from the subjective point of view, making sense of our own actions in choosing them, we must understand those actions in terms of law-like relations. Just as it is a requirement of our knowledge of the physical world that we have knowledge of the laws that govern that world, so it is a requirement of our knowledge of the moral world that we seek out both subjective and objective understanding of moral activity within the context of law-like relations. Our understanding of morality as motivated agency is therefore defined by the agents' articulation of maxims, just as our understanding of the universe is defined by physical laws. Therefore, to understand the moral significance of any action we must know both the means that the agent chooses to advance his ends and the nature of those ends themselves.

---

[4] It is perhaps worth noting that our conception of the laws of nature (including certain probabilistic effects of quantum level phenomena) is rather different than Kant's Newtonian presuppositions. Nonetheless, the nature of his thinking should be clear enough.

Subjectively, i.e., when we are in the process of deliberating about what we are to do, we must articulate our practical reasoning in the form of a law by formulating a maxim to describe both our desired ends and our plan for achieving those ends. However, Kant thinks a problem arises so long as we concentrate only on the means side of our practical reasoning, taking our ends as given (by desire, for example). In doing this we are merely formulating hypothetical imperatives, conditionals that describe the means to achieve some given end, but which do not prescribe the end itself. Thus, in hypothetical imperatives reason is used merely instrumentally, in the service of ends which it does not evaluate.

To the extent that those ends are set by the agent merely in response to his desires, according to Kant, the agent acts *heteronomously* with respect to his will. Such desires are not the product of his rational deliberation on what ends are most appropriate, but brute facts of his psychology, possessing no independent moral worth. Kant thinks that so long as we make sense of our practical reasoning through hypothetical imperatives alone we are unable to achieve a fully rational ground for our practical deliberations. The kind of unchosen desires which inform such hypothetical imperatives as ends are without independent moral status simply because they are not the product of any kind of rational deliberation. They come upon the agent continually and none able to justify itself beyond the brutal fact of a *want*.

As I have already noted, hypothetical imperatives are conditionals; they have the form: (in conditions C) in order to achieve *end(s)* ψ, I will do *action(s)* ϕ. They do not however answer the question: but why do I want ψ? So long as you cannot give a rationally self-sufficient justification for your end, Kant thinks your practical reasoning will be caught in a regress. Hypothetical

imperatives will only give rise to further hypothetical imperatives, without an ultimate *reason* that can ground the entire process of practical reasoning.[5] Thus the satisfaction of one desire only leads to others, each similarly insistent, each similarly unjustified. This, then, is the significance of the Categorical Imperative; the Categorical Imperative is supposed to be the self-sufficient ground of all (practical)[6] reason. It provides a way out of the regress of cascading hypothetical imperatives by giving a reason for acting that requires no further justification. And this reason for acting just is what rationality itself requires.

The humanity formulation of the Categorical Imperative (one of three, according to Kant, logically equivalent formulations of the CI) makes this clear by requiring us to recognize all other rational beings as ends in themselves.[7] When Kant claims that persons are ends, what does 'ends' mean? The second formulation of the Categorical Imperative clearly opposes *means* to *ends*. The requirement to treat humanity as an end suggests that I am not to treat persons as instruments for achieving some other purpose or goal of my own. However, what would it mean to treat a person *as* a purpose or goal? It seems much more natural to speak of a person *having* a purpose or goal, not *being* a

---

[5] It is true that one's deliberations as to ends might lead simply end with "because I want it" or "because it gives me pleasure." However, for Kant, such "reasons" are not real reasons at all; they possess no self-sufficient justification, and so merely reflect a will that is *acted upon* (by non-rational desires, etc.), rather than a will that acts (i.e., can give real *reasons* for its actions).

[6] Again, O'Neill would leave out the 'practical' and claim that the CI is the ground of all reason. I choose to assert only the weaker claim in Kant's name, confident that if O'Neill is right about the superset, then I am not asserting a falsehood by predicating the same property of the subset (leaving the status of the disjunction open).

[7] The three formulations are: 1) The formula of the universal law: "Act only according to that maxim whereby you can at the same time will that it should become a universal law" (421); 2) the formula of humanity: "Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means" (429); 3) the formula of autonomy: "Always choose in such a way that in the same volition the maxims of choice are at the same time present as universal law" (440).

purpose or goal. So what does Kant mean when he claims that persons are "ends in themselves"?

An *end*, according to Kant, "has in itself an absolute worth" and alone can be "a ground of determinate laws" (428). "[R]ational nature exists as an end in itself;" it has no further purpose beyond itself. When Kant talks of persons as ends, what he really has in mind is that persons are fundamentally *causes*. Persons are ends in themselves because, for the purposes of moral discourse, a person (and only a person) can be the self-sufficient end (i.e., beginning) of the causal chain whose effects are the sorts of actions we conventionally subject to moral evaluation.

In other words, to be a genuine agent, to bear responsibility for his acts, a person must be the free cause of those acts. With regard to the moral actions for which he is responsible, the agent must be a kind of Prime Mover, and it is only insofar as he is the first cause that he himself *acts*. Persons are ends in themselves insofar as they have the capacity to be free causes. This proposition is at the heart of Kant's understanding of pure morality and accounts for the central importance of autonomy in his formal ethics. He begins section III of the *Groundwork* by claiming that the "will is a kind of causality belonging to living beings insofar as they are rational; freedom would be the property of this causality that makes it effective independent of any determination by alien causes" (446).

It is in virtue of being ends in themselves that persons are "objects of respect." Respect is a necessary logical requirement of recognizing what persons most fundamentally are—i.e., free causes. It is this recognition that grounds the demand for respect because it is most fundamentally a recognition of another's rationality—another's capacity to be a self-existent

end. This is why the second formulation notes that one must treat humanity, "whether in your *own* person or in the person of another" (emphasis added), as an end. Each "man necessarily thinks of his own existence" as a self-sufficient end, but so "also does every other rational being think of his existence on the same rational ground" (429). Thus, in recognizing myself as a self-sufficient end I must at the same time necessarily recognize every other rational being as a self-sufficient end. In other words, I cannot rationally conceive of myself as a free cause—as the ultimate source of my own activity—without conceiving of all other creatures with the same rational capacity in the same way. Another way to see the point: to fail to recognize the ultimate value of another rational being would constitute a failure of my own rationality. It would be intrinsically irrational, much like denying a truth of mathematics, and would represent a fundamental misunderstanding of what is at issue.

Respect and autonomy are, in some sense, the flip sides of the same coin. The will that is a free cause, in order to value its own autonomy properly, must value all other rational creatures in the same way by paying them respect. This is an absolutely key point, one missing from many discussions of respect for persons. This is why Kant thinks that respect comes out of the nature of practical reason itself; it is entailed as a logical consequence of the realization by each agent that he is a free cause.[8]

[8] I hope it is clear that here (and elsewhere where I am using quasi-perceptual verbs like 'recognize') I am not making a point merely about an agent's subjective psychology. An agent is a free cause whether he realizes it or not; after all, Kant thinks agents were free causes before he came along to do his transcendental deduction. The point is merely that Kant thinks that he has shown how agents must think of themselves in order to have an adequate understanding of themselves *qua* agents.

Within this context then we have a fully grounded reason for a demand for *equal* respect. For a rational creature to cognize his own rational nature he must understand every other rational creature to possess the same nature he has. To disrespect another rational creature is to undermine one's own rationality; thus, reason itself requires respect because respect just is the practice of reason with regard to other instantiations of itself.

What this recognition actually entails is that my own status as a moral creature—responsible for my moral actions as a free rational cause—requires that my moral life play itself out within the space of moral reasons. Accordingly, anyone else who is also within that space of moral reasons is condign of the same regard I pay myself when I recognize myself as a moral agent. Thus, I cannot treat another as a means without in essence failing to understand *myself* as a locus of causation within the moral world. To treat another as a means is to abdicate my own rationality. To fail to pay respect to other rational creatures is to fail to see myself as an ultimate cause, and thus as a moral creature.

Within this context, then, we actually do have a fully grounded reason for the demand for respect. Recall that it is supposed to be an analytic requirement of practical reasoning to understand that practical reasoning must occur within law-like formulizations—i.e., maxims. This is because agents have to serve as primary causes of their actions in order to be morally responsible for them. As such, the nature of practical reasoning itself compels the recognition of all who engage in law-governed practical reasoning *as* agents. The price of admission to the sphere of moral reasons is precisely the recognition of all fellow practical reasoners as ultimate moral causes. Thus, for Kant, respect is an *analytic* requirement of practical reasoning.

So far we have seen how the nature of practical reason constrains agents to respect the agency of others and so treat them always as ends in themselves, and not as mere means to one's own ends. There is an important loose end, however, that now needs to be taken up. I mentioned above that being motivated merely by one's unreflective desires is to be *heteronomous* with respect to one's own will.

The will is heteronomous when it "does not give itself the law, but [its] object does so because of its relation to the will" (441). To be heteronomous with respect to one's will is to be passive before one's own desires; it is to suffer a passion, rather than to choose one's own ends. A will that is heteronomous fails to achieve true autonomy and is thus fundamentally unfree. "The moral imperative must therefore abstract from every object to such an extant that no object has any influence at all on the will, so that practical reason (the will) may not merely minister to an interest not belonging to it but may merely show its own commanding authority as the supreme legislation."

Thus to be *fully* moral, a will must be abstracted from its objects (e.g., what one desires) and focused on giving itself the law.[9] An autonomous (moral) will would "always choose in such a way that in the same volition the maxims of the choice are at the same time present as universal law" (440). It will of course surprise no one that this willing as universal law just is the Categorical Imperative.

[9] "For the proper and inestimable worth of an absolutely good will consists precisely in the fact that the principle of action is free of all influences from contingent grounds, which only experience can furnish" (426).

The form of the universal law requires simply that it be exceptionless and necessary, as Kant conceived of the laws of nature to be.[10] Thus, for a maxim to be "at the same time present as a universal law" "in the same volition" means that the maxim must be *willed* as exceptionless and necessary. Which is to say, the maxim must be willed as binding on all (rational) agents (in the same circumstances). "The concept of every rational being as one who must regard himself legislating universal law by all his will's maxims… leads to another very fruitful concept… a kingdom of ends… [which is] a systematic union of different rational beings through common laws" (433). The kingdom of ends then is "a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and also the particular ends which each may set for himself)."

This principle supplies the content of Kant's moral philosophy. As I noted in passing in the last chapter, Kant's chief, indeed only, concern is the nature of the agent's will. "There is no possibility of thinking of anything at all in the world, or even out of it, which can be regarded as good without qualification, except a *good will*" (393). Thus, "what is essentially good in the action consists in the mental disposition, let the consequences be what they may" (416). Therefore, what matters for Kant is not what one does, but *why* one does what one does. The moral worth of any action can only be determined by an examination of the motives of the agent.

Of course, as we have already noted, one problem with depending on motives in evaluating the moral worth of an agent's actions is that motives are often obscured, sometimes even from the agent himself. Thus Kant's project is

[10] It is of course no accident that universality and necessity are the "marks" of the *a priori* for Kant.

to describe the binding precepts of morality in terms of what a good will must will in order to be good. Whether any given person has such a will is unknowable to human beings, so much so that we cannot know if any action has ever had true moral worth. People have certainly acted in accord with the demands of the Moral Law, but often these actions have not been done out of a sheer regard for duty, but because of some natural inclination.[11]

In other words, if I have told the truth to my mother, not because it is immoral to lie, but because I love her and so cannot bring myself to deceive her, then my actions—though in accord with the demands of morality—do not themselves have moral worth. In order to be morally worthwhile, my actions have to be motivated by a respect for my duty, and no "contingent grounds" such as my love for this particular woman, can influence my will. Kant realizes that this is an amazingly stringent demand and is prepared to acknowledge that it may deprive of moral status almost all of the actions we might pre-reflectively have believed to be moral, since the motives behind those actions were not a pure respect for the moral law alone.

One of Kant's most famous examples illustrates this point. Kant argues that the charitable actions of a misanthrope are more likely to have genuine moral worth than the same sort of actions done by a hearty humanitarian who takes pleasure from his charity. It is precisely because the misanthrope finds other people disagreeable, and so acts solely out of a sense of duty, that his charitable actions have moral worth. The humanitarian's similar actions, while in accord with what morality demands, do not have independent moral worth

[11] At 397 Kant discusses the case of the honest shopkeeper who is honest only so has to maintain his reputation and so secure future customers. His actions—though in accord with what morality demands—do not have moral worth because they are not done out of a respect for duty.

because the motivation behind them is infected by the pleasure the humanitarian takes (and expects to take) from his charity.

It is most important for us to see, in Kant's framework, that morality, and indeed the requirements of duty generally, bind agents only *qua* practical reasoner. I am duty-bound by nothing other than the nature of my own rationality, and thus my actions are morally worthwhile only insofar as they proceed solely from a respect for my duty. Thus, the key feature of Kant's *moral* theory is purity of motive, of acting out of a recognition of the demands of the Categorical Imperative, which is acting always such that the maxim describing my rationale for acting could be universalized for all rational agents.

This point can generate some confusion. It is not the case, of course, that an action in accord with the demands of morality that is done from a private (i.e., contingent) motivation is necessarily *im*moral. The point is that such an action is not *morally* worthy—i.e., has no independent moral value. Such an action is not a reflection of a "perfectly good will." This then would seem to generate a *prima facie* duty to cultivate a will which will not be determined by such "contingent grounds" as affection or friendship.

It is at the point that many critics of Kant, myself included, wish to attack his theory, and I will have more to say about such criticisms below, but it is important that we be clear about what is at stake here. Kant is not committed to the ridiculous position that being moral requires one to have no special relationships; rather, he is only committed to the proposition that the *perfectly good* will cannot be determined by desires or motivations arising from those relationships. Now it may be the case, and Kant realizes this of course, that the limitations of human nature—the fact that we are *imperfectly*

rational—precludes the possession by any actual mere human being of a perfectly good will. Kant can accept this without a problem for his moral theory so long as he acknowledges that the motivations, demands, and desires arising from special relationships are simply outside the province of morality.

Nonetheless, this is an extremely important consequence of Kant's moral theory and clearly important to the question of what Kant can say about the demands of love and loyalty. What we need to see is that this consequence follows inescapably from the nature of Kant's project. By constraining morality to just those demands that bind all rational actors, in virtue of their rationality, he leaves no place for sentiment of any sort to enter into the deliberations of practical reason with independent moral force. Thus, within Kant's moral system, the demands of love and loyalty cannot have any genuine *moral* value insofar as that concern is founded in the specificity of sentiment towards particulars.

Again, it is worth emphasizing that this conclusion does not necessarily rule out responding to such demands. However, so far as one understands oneself to be motivated by a demand of love and loyalty, one's actions lack moral worth. Thus for instance, if I tell my mother the truth just out of the love I bear her (so I cannot bring myself to lie to her)[12] my action has no moral worth. Obviously, however, that does not make it immoral. When I tell the nice lady next door the same thing out of a respect for my duty to tell the

[12] The experience of this kind of inability I will describe in the next chapter as a species of the phenomenon that Frankfurt calls "volitional necessity." Now, this obviously does not mean that in no situation would I bring myself to lie to her. If I truly believed that deceiving her would (legitimately) shield her from some great pain, for instance, the same kind of volitional necessity might manifest itself in a decision to deceive her. On the other hand, Kant can leave no room for this, at least within the sphere of morality. To lie to her for any reason, even "for her own good" would be to disrespect her dignity as a moral agent (and so, *ipso facto* it could not have been for her own good).

truth, the same action is morally worthy. Thus, I have *done* nothing *per se* wrong in responding to the warm demands of filial love; however, I have also done nothing morally worthy.

So far I have taken pains to make a kind of *amoral* space for the demands of love and loyalty. Being motivated by these demands may, and often will, generate actions that are objectively legitimate, i.e., actions that are indistinguishable from those that would be performed by an agent with a perfectly good will in the same moral choice situation—just as my telling my mother the truth was demanded by morality, even though I actually did it out of my love for her. However, insofar as the Categorical Imperative calls us to justify all of our actions by a willingness to will the maxim embodying them as universal and necessary, these actions are subjectively illegitimate. They express an imperfectly good will.

I think it an open and inadequately explored question exactly how much of a problem this need be to Kant. One might think that he could take refuge in the simple fact that human beings are *imperfectly* rational creatures, and so never truly in a position to possess a perfectly good will (unlike God and His angels). Kant says in another context "out of such crooked wood as man is made, nothing straight can be built"; I am inclined to think that he is sympathetic to this view of the intrinsic imperfectability of men's wills.[13]

Whatever one might think of how serious a problem this is for Kant—or indeed if it is a problem at all—it should be clear that the demands of love and loyalty have no place within Kant's system as ultimate *reasons* for acting.

---

[13] This famous quote appears in sixth proposition of his essay "Idea Towards a General History (from) a Cosmopolitan Outlook." This kind of outlook, of course, would be quite consistent with Kant's rather strict Lutheran family background. Other readers of Kant seem to find this line less appealing, and so are apt to generate *apologiai* of various sorts on this issue.

Consider a maxim such as: In most cases I will favor my family and friends over strangers in order to express my love for them. A will acting on this maxim would be heteronomous because the maxim expresses an end that itself is "contingent."[14]

Again we need to be clear: to act on this maxim is not necessarily to act immorally. That would only clearly be the case if the maxim could not be *willed* as a universal law. For Kant, in testing a maxim we need to know both whether it is consistent—i.e., can be *conceived* as a universal law—and also whether it can be *willed* as a universal law.

About the maxim in question, we can see that if everyone were required to favor his family, then each agent would still be able to do so. In a world where everyone must favor his own family, it is certainly possible to favor one's own family—so the conception test is passed. The second question is whether we can *will* such a world. It does not seem to me that this would be a problem. We could not will a world in which people are positively prevented from helping people who are not members of their families. However, the maxim in question—to favor one's family members—does not absolutely preclude helping others above the favor one renders to one's family. In other words, I can will such a world without undermining the ground of my own rationality.

There may be limitations on the extent to which I can favor my family. I think it would probably be impossible to will a world in which I am required to favor my friends and family above all others in all cases. For example, it would be ridiculous to will a world in which one would be unable to tend first

[14] Of course it is not transparent what it would mean for me to take as my end the expression of love. Nonetheless, if it is founded on sentiment and affection it does seems clear that Kant would consider it contingent.

to a stranger with life-threatening injuries who happened to be in the same car accident in which one's beloved child received a skinned elbow. For to do so would be to will a world in which I preclude strangers from helping me, thereby irrationally threatening my own rescue in cases of accidental debilitation.[15] Thus this kind of Kantian moral deliberation would certainly set permissibility limits as to the situations in which favoritism is rationally willable, but it would nonetheless leave some space for such favoritism.

What this reveals is that taking such a maxim is weakly permitted—i.e., it does not directly violate the requirements of rational morality for imperfectly rational creatures. What it does not do, however, is provide any grounds of thinking that morality *requires* anything like the end of this maxim. To see exactly why this is we need to remember how Kant's moral system is supposed to make its claims upon us—viz. solely in virtue of our rational nature.

This does not mean that Kant was unaware of, or ignored, the contingent conditions within which our rationality must play itself out.[16] As I suggested above I think Kant was quite alive to the limitations inherent in human nature, and so our tendency to be always morally lacking. However,

---

[15] The reason for the contradiction here is the same as the fourth example of non-beneficence (423). As we are dependent, needful creatures, we will always require the potential assistance of others. Accordingly, we cannot will a world in which such assistance is precluded to/from persons to whom one does not already have a special relationship. This may be slightly too strong; it is imaginable I suppose, that some humans have lived their entire lives without ever being among others with whom they did not have a special relationship (e.g., a very primitive tribe based on kinship). In such conditions it is perhaps imaginable that one could will a world that depends on special relationships alone to motivate others to help.

[16] The fourth example of non-beneficence I mentioned before only makes sense to the extent that our dependence on others is foreseeable as necessary for the maintenance of our rationality. Of course, this raises another question for Kant; namely, would beneficence be morally required of rational creatures who did not suffer the needs of interdependence as we do? If my fellow creatures were never in a position to help me, or if I were never in a position to require their help, would it really be impossible to will a world in which that help could not be proffered? I do not see that it would.

although those contingencies inform and structure the conditions within which rational choices have to be made, the actual foundation of all moral obligation—the ground of the *ought*—lies solely within our capacity for rational deliberation.[17] As such, there is simply no room in Kant's moral philosophy for a legitimate obligation—a legitimate *ought*—to be grounded in love or loyalty.

This is exactly where we need to be most careful, and where there is the greatest potential for confusion for both Kant's critics and defenders. What is at stake here is the nature of the complicated interaction between the wholly rational ground of morality and the contingent circumstances and contexts which provide both the conditions in which real practical deliberations must be made and the objects of our non-rational wants. This is where, according to one of Kant's more sophisticated modern defenders—Barbara Herman—*autonomy* (the ground of morality) interacts with *agency* (the exercise of rationality in the world). At this point I would like to turn to her discussion to see how she thinks these complicated interactions might play out in such a way that there is "space for the claim that connection itself could be partially

[17] In fact, the ought-ness of the ought, so to speak, is a result of our capacity to do evil. In other words, the fact that moral obligations present themselves to us as an *ought* rather than the automatic determination of the will (as with perfectly rational creatures like angels, or the most perfectly rational being—God) shows how it is necessary for us to integrate the demands of rational nature into the context of our wills' many contingent objects. It is precisely because we are tempted to do wrong that the demands of morality are recognized as an "ought to do," rather than a mere "will do." If we were perfectly rational creatures the demands of morality would not really be demands at all; our wills would turn to them naturally and without possibility of contradiction. Thus the very fact that morality is a struggle for us—a struggle against heteronomy in Kant's terms—shows that Kant's thinking makes room for the contingent conditions of human existence. However, what his thinking refuses to do is allow these contingencies to infect the foundation of morality. The fact that the demands of morality are informed by the contingent conditions of the world in which humans find themselves does not mean that those demands are in any way dependent on those conditions. In differing conditions, the demands might have different forms or different objects, but the ground or source of those demands would remain always the Categorical Imperative alone.

dependent on or a function of moral value" (Herman [1991], 780) within the Kantian framework.

Herman

Herman explicitly addresses herself to the (potential) alienation felt by agents in response to the demands of a universalist (i.e., impartial) moral theory. She takes up the torch for the Kantian position against the kinds of complaints we saw in previous chapters from Bernard Williams. Herman thinks that this kind of criticism, that Kant's moral theory cannot adequately account for the demands of "attachment," is predicated on a mistaken picture of the nature of practical deliberation. She thus offers, in place of the model she takes Kant's critics to be using (the "plural interest model"), an alternative (the "deliberative field model"), which she claims will make room for the demands of love and loyalty within a picture of integrated, mature moral "agency."

The problem is that the demands of love and loyalty (Herman refers to "attachments") often seem to conflict with the demands of impartial morality. This can lead an agent who cares about morality to feel split or conflicted, unable to respond to one set of demands without offending the other. Especially insofar as impartial morality demands that the agent give up relationships that are centrally important to him, this can become a burden that threatens to undermine his very commitment to morality. Herman claims that the picture of practical deliberation these concerns presuppose is the "plural interest model."

> According to the first or plural interest model, where there is connection, there are those I care about, and the effect of my caring is to give their interests greater deliberative weight: for me. They matter more. And they matter more to me because I care about them. When I need to balance or weigh interests—should I do some good for my son or his friend—my son counts more…
>
> On the plural interest model, when morality contends with attachments it forces one against the grain, attacking the immediacy of connection. It would be natural to feel hostile to or alienated from the requirements of morality if they in this way denied a deeply felt claim of partiality.... The problem arises when it looks like "over here" is what I most care about, what I want to happen (and cannot not want to happen), but "over there" is what impartial morality demands. There is then deep conflict and tension. And when impartial morality wins, it is not only at the expense of what I most care about, it provides no deliberative space even to acknowledge my concerns. (Herman [1991], 782-3)

This view's proponents see practical deliberation as a kind of empty space into which flow various demands—both the demands of love and loyalty and the demands of impartial morality (and, perhaps, demands of all sorts, from simple bodily desires to complex aesthetic ones, etc.). The purpose of deliberative reason is to weigh these various demands against each other and to determine the course of action that best satisfies the demands of what I care about most. If in a given case the demands of love and loyalty conflict with the dictates of morality, the result is a feeling of alienation or conflictedness before these irreconcilable forces.

According to Herman, this model suffers from a major defect—a kind of fundamental immaturity. This immaturity consists in a failure to fold together both the demands of love and loyalty and impartial morality into a fully human life that is informed by both. "Among the elements of a full moral theory we should find an account of how one is to integrate the requirements

of morality into one's life." Accordingly, she offers a different, better model of practical reason—what she calls the "deliberative field model."

> According to this deliberative field model, the practical self does not have as its major task negotiating a settlement among independent competing claims. Insofar as one has interests and commitments, one is a (human) self. But a human life is not the resultant "bundle" of competing interests (among which is an interest in morality) One's interest[s] are present on a deliberative field that contains everything that gives one reasons. Thus, in addition to interests and attachments, there are also grounds of obligation, principles of prudential rationality, and depending on the individual, a more or less complex conception of the Good. (784)

According to Herman the key advantage of this model is that it allows for "the integration and transformation of the ends in light of one another, of one's practical situation, and of one's conception of place and importance understood the through regulative principles—aesthetic, moral, prudential—one accepts" (785-6). This process of integration involves bringing the demands of love and loyalty or attachment within a single deliberative field by "normalizing" them "to varying degrees to the principles of practical agency, both moral and nonmoral" (789).

The idea is that the demands of love and loyalty, or anything else, can only have a place within the sphere of practical reason—that is, can only appear within the scope of rational practical deliberation—once they have been processed into an intelligible form and thereby become subject to judgments. As Herman says, desires "do not give reasons for action: they may explain why such and such is a reason for action, or even why something can be an effective reason for action, but the desire itself is not a reason. One can

take the fact of a desire to be a reason, but that is just to hold that desire, or this desire, is good" (785).

Consider a simple desire such as the desire for chocolate cake. By itself, as a psychological phenomenon, that desire does not supply a reason for acting. Rather, it can become a *reason* only insofar as I form a judgment. I could form a simple syllogism by first forming a maxim: I will obtain and eat chocolate cake in order to satisfy my desire for chocolate cake. I judge that it is good for me to satisfy a desire for chocolate cake. I conclude that it is good for me to obtain and eat chocolate cake. Thus, I have a reason to do so.

For Herman, a desire is important (i.e., can enter into the deliberations of practical reason) only when conjoined to a judgment that fulfilling that desire is good. Her concern does not extend beyond reasons, thus for a desire to play any role whatsoever in her moral psychology it has to be "normalized" in this way and then integrated into the all-encompassing deliberation in which various reasons are considered in light of each other.

Implicitly, however, this ends up representing action as proceeding merely from practical reason. For Herman, all that matters is *deliberation*—i.e., practical reasoning. When she addresses herself to the challenge that deep-seated attachments represent to impartial morality, she sees this challenge entirely through the lens of practical rationality. These deep-seated attachments make themselves felt only as reason giving judgments within the "deliberative field" (i.e., that these attachments are good). The absolutely key thing to see is that her deliberative field—or the space of practical reason, or whatever else it might be called—presupposes the primacy of reasons and reason-giving. Attachments—the demands of love and loyalty—matter only insofar as they can generate reasons through judgments that they are good.

For Herman, the field itself, to stretch her metaphor, is flat. All psychic demands (be they the demands of impartial morality or of love and loyalty) enter the field through the normalization process—that is, they are or become reasons. These demands play themselves out on that field according to the "rules" of the game—what she calls "the principles of practical agency, both moral and nonmoral." These rules define the game by defining the players—through judgments yielding reasons (which in turn are defined by principles that describe what counts as a well-formed reason). The ultimate goal of the moral theorist is to describe these rules, and so ultimately define the game and the "victory" conditions (i.e., what action is best all things considered).

It is certainly the case that under Herman's model the demands of love and loyalty can generate reasons and so enter into her "deliberative field." My point is that it is not in generating reasons that the *deepest* commitments affect the will. As I will argue in the next chapter, on an alternative (and more plausible) account of the will such demands can directly impact the will, because they are partially constitutive of it. On this picture of the will, it is larger than practical reason, and so deep-seated desires can affect the will directly without becoming reasons *per se*.

I think Herman fails to account for the fact that the deepest attachments of human life flow out of volitionally constitutive core commitments and do not enter into the deliberative field as legitimate players (i.e., as reasons). Instead, they tilt the field itself. They are not judged; there is no judgment that they are good. They do not give reasons. Rather by shaping the will directly (i.e., tilting the field) they prescribe the very limits of the game. Far from being normalized by the principles of practical agency, they can prescribe the very scope of those principles, by setting the limits within which deliberation can

take place. They do not primarily affect the will by giving reasons, but define the horizons within which reasons can matter. In other words, their presence in human life defies the notion that reasons are *always* basic.

It is here that we have finally reached the root of the disagreement. Kant and Herman, et al., are claiming that the very structure of practical rationality places real restrictions on what can count as a *reason*. On this view, since practical rationality is constitutive of moral agency, these structural features are the inescapable conditions of moral action. Once we see this we are supposed to understand that the demands of love and loyalty only have standing within deliberation to the extent that they can be made to participate within the structure of reasons. This means that whatever real moral content they have will need to be formulated in the terms of maxims to even enter into consideration. And since the possible permutations of consistent maxims are constrained by the nature of practical reason, so the scope of possible *justified* action is constrained.

Fascinating though this notion is, I think it utterly misses the point of Williams' example. Kant (according to Herman) has described the limits of the deliberative field by showing how practical reason requires its potential objects to be normalized within the formal structure of maxims. Even if this were true, however, this would not capture the deep intuition to which Williams appeals in the case of the wife in peril.

The point of Williams' example is not that the demand of love which motivates the man to save his wife can be justified or "integrated" within the deliberative field. The point of that example precisely is that deliberation is here inappropriate—this kind of deliberation is exactly "one thought too

many." What is at stake in the deepest attachments is not justification within practical reason, but their capacity to undermine it altogether.

The disagreement between Kant's critics and defenders is very deep here, because the disagreement is a disagreement about the very nature of the will. For Kant, the will just *is* the faculty of practical reason. For Williams and others like Frankfurt, the will is in a sense more complicated, because practical reason is a *constituent* of, but not entirely *constitutive* of the will. Thus the structure of the will cannot be fully described only in terms of practical reason. The movements of the will need not always take the form of maxims describing valid means-end reasoning; sometimes, in cases like being moved to save a beloved wife, the will itself will be directly determined by the centrality of "what we care about" most (Frankfurt), those things which constitute the "ground projects" of our lives (Williams).

There is a very important distinction to be drawn, inattention to which can produce a great deal of confusion on these issues. This distinction concerns the nature of justification itself as applied to these kinds of cases. What we need to understand is that the kind of "justification" on offer here is necessarily external to the deliberations of practical reason. Indeed it is perhaps misleading to speak of *justification* at all, in these sorts of cases.[18]

The reason for this is that, as Kant points out and Onora O'Neill is especially keen to emphasize,[19] the structure of reason—that is, simply, the

[18] To see this, we need only consider the question: is some unalterable feature of human nature justified? The question seems almost nonsensical, analogous to asking whether our having four fingers (instead of three, say) is justified. Even if this were granted this would be a strange kind of "justification." In Part II of my dissertation, I basically argue that to be fully human, we must possess the kind of will that directly expresses the values we try to classify as demand of love and loyalty. Some of this argument is rehearsed below, but in a very truncated form.

[19] See especially Chapter Three, "Reason and autonomy in *Grundlegung* III", in O'Neill [1989].

giving of reasons—constrains and informs the possibilities of deliberation. Accordingly, on a Kantian view justification is necessarily limited to only those things that can be expressed in its own terms—in the terms of maxims. However, the suggestion that I think Williams is making is that there is a kind of determination of the will that bypasses, or more accurately *undercuts* the processes of deliberation. As more basic than rational deliberation, these determinations cannot be *justified* in merely the conventional way, precisely because they are the ground of the will itself. This is not to say that they cannot be justified at all, but such a justification would require an altogether different sort of argument.

When we turn our attention to a case like the one of the man saving his wife, we consider it not only after the fact, but also, more importantly, in tranquility and divorced from the immediacy of the demand on his will that his love for her generates. In other words, the search for justification of the conventional sort is almost doomed to failure from the start, because the kind of "justification" on offer cannot be properly formulated from a *deliberative* perspective. The justification is nothing more than this: he could not have *willed* it otherwise. Not because he introduced a maxim describing his means and ends into his deliberative field and judged them consistent with the demands of the Categorical Imperative, but because his will was already determined in this direction prior (conceptually as well as temporally) to his deliberating. In other words, some of the demands of love and loyalty simply cannot be "normalized" as Herman would like, precisely because they operate outside the space of reasons, and thus outside the "principles of practical agency" *as she understands them*.

This is not to say that no reasons might be given for them, however. The reasons given, though, will necessarily be external reasons, reasons having to do with brute facts about the kind of creatures we are and the necessary structure of our wills. The justification, in other words, will be simply that it could not have been otherwise, given the kind of creatures we are.

Now we are brought face to face with just how alien these demands of love and loyalty can be to deliberative reason. I want to be clear; it is not the case that the demands of love and loyalty cannot generate "normalized" reasons as well. Indeed in many cases they can and do; however, at their deepest level they operate "behind" or "under" deliberative reason altogether. They ground the will, by giving it determinate objects of central concern that it does not choose.[20]

These points will be taken up, rearticulated, expanded and defended in Part II. At this point my main concern is to show how far the Kantian project is from actually considering the demands of love and loyalty as expressions of the deepest values in human life. Kant's project, from the very start, is by its nature hostile to the demands of love and loyalty, because it simply cannot consider them without fundamentally altering the character of the values they express. As such, to the extent that these values are "normalized" they are eviscerated, and it becomes no wonder that they fail to find a satisfactory justification within such a system.

[20] This is not to deny the possibility that a disordered will might not very well attempt, and to a large degree succeed, at destroying its own ground through a conscious nihilism.

Heteronomy Revisited

Of course, it is likely that Kant (and Herman) would respond to my claims about the love and loyalty undercutting deliberative reason by observing that even if this is how the demands of love and loyalty work upon the will, any agent whose will is determined according to these demands merely suffers from heteronomy of the will. Recall that according to Kant, the will is heteronomous when it "does not give itself the law, but [its] object does so because of its relation to the will" (441).

The obvious question is how seriously we should take Kant's notion that autonomy just is the abstraction from every "interest not belonging to" the will such that it "show its own commanding authority as the supreme legislation" (441). Specifically, do we not have reason to think that Kant is begging a very important question by assuming that the demands of love and loyalty do not truly *belong* to the will? In other words, to make sense of Kant's understanding of autonomy, do we not just have to *assume* that the will is merely practical rationality itself?

For Kant, the demands of love and loyalty seem to be properties of their various objects rather than properties of the subject (i.e., the agent). As he apparently conceived of them, these demands act *upon* the will from outside. To understand his distinction between heteronomy and autonomy, there must be a clear difference between what is internal and what is external to the will, and the demands of love and loyalty must fall outside of it.

I suspect that much of the intuitive pull of the Kantian notion of autonomy lies in the apparently intimate connection between moral responsibility and choice. When we are acted upon by very strong passions, it

is as if the faculty of choice is substantially undermined so as to mitigate moral responsibility itself.[21] In order to make sense of this, however, we have to understand these kinds of passions as external to our real selves—our free moral agency.

Even if this is the case with passions like consuming anger, however, it is not at all clear that it is appropriate to consider the demands of love and loyalty in this way. Structurally speaking, the demands of love and loyalty are not "objects" that act upon the will. They do not act *upon* the will from outside of it, but are expressions *of* the will itself. The deepest demands of love and loyalty manifest themselves as what Frankfurt calls "volitional necessities," and so proceed from and reflect the internal structure of the will.[22]

Thus, the demands of love and loyalty should not be understood to threaten autonomy, because they are not "objects" of the will at all, but rather products of it. They cannot threaten the will's freedom, because they themselves are the ineluctable consequence of having a human-shaped will at all, expressing the agent's central volitional commitments, which are themselves (partially) constitutive of that will.

As will I will argue further in the next chapter, it is a mistake to conceive of the free will as a will completely unencumbered by volitional commitments. For a choice to constitute a real choice—an expression of the volition of an agent—it must reflect the deepest commitments of that agent himself. In other words, moral choice—a choice legitimately subject to praise or blame—must be the expression of a stable will, a will constituted by

[21] Thus, we have the intuition that a "crime of passion" is somehow less blameworthy than a premeditated one. This is the moral underpinning of the temporary insanity defense, as well, but this topic is sufficiently convoluted to be well beyond the scope of our present discussion.
[22] This notion from Frankfurt is discussed and developed at length in the next chapter.

commitments to what the agent loves. Otherwise, the "choice" will be either arbitrary or impersonal.

Autonomy conceived of as Kant conceived of it is radically impersonal. The execution of the demands of the Moral Law done out of a respect for that law (i.e., the only morally praiseworthy actions according to Kant) must necessarily be what any other rational agent should do in the same choice situation. The Kantian metaphor of the "Kingdom of Ends" captures this perfectly—in correctly giving himself the law, every agent also gives the same law to all other relevantly similar agents. Nothing signifies the essential uniqueness of any given agent, because *qua* moral agent, all agents are interchangeable.[23]

I think it is a mistake to conceive of autonomy simply in terms of what all moral agents fully share; rather it is autonomy is best expressed in terms of what makes each of them unique. The demands of love and loyalty thus do not undermine the agent's autonomy, but are an expression of part of what fundamentally grounds it. The confusion could only arise because Kant and his defenders beg the essential question of the relationship of practical reason to the will. The Kantians set up the question in such a way as to presuppose that autonomy proceeds entirely from the exercise of practical reason, and in so doing, they presuppose a moral psychology that does terminal violence to

[23] For Kant, all moral agents, in order to be moral agents, must fully share the formal capacity for practical reason. They obviously do not all have to have the same actual capacity. Furthermore, differences in the situations in which that rationality plays itself out may generate certain differences between significantly different classes of moral agents. In the famous fourth example of non-beneficence (in the *Groundwork*), contingent facts like our corporeality in fragile bodies have important consequences for understanding our moral duties, which may not affect other rational moral agents differently circumstanced (e.g., unembodied rational intelligences). See Herman's discussion in "Mutual Aid and Respect for Persons" (Herman [1993]).

some of our strongest intuitions regarding the importance of attachment, loyalty, and love.

Ultimately the strong intuitions that underlie our valuing of autonomy and disvaluing of heteronomy are a reflection of the fact that for an agent's choices to be free, they must be *his* choices and not externally compelled. The agent must "give himself" the volition. The mistake in the Kantian picture lies in the undefended (and indefensible) presupposition that only pure reason itself authentically belongs to the agent.

In the next chapter we will see instead that when acting on the deepest demands of love and loyalty, which reflect the internal structure of his own will, the agent's acts are as fully authentic as any acts can be. To act through a volitional necessity is to make manifest the deepest structure of the will, it is indeed the purest expression of that which makes the agent the agent that he is. Nothing, especially not acting from mere respect for the Moral Law, could be more authentic, more *self* governing.

The Formula of Humanity

Thus far I have concentrated on what I take to be the most innovative aspect of Kant's moral philosophy, namely his claim to have described, in the categorical imperative, the fundamental principle of pure (practical) reason. Ultimately, I have claimed, this leads to the collapse of the distinction between the will (agency) and practical reason. The perfect will, as agency (i.e., the autonomous actor acting in the world), is precisely and completely described in rational terms. On this model, the demands of love and loyalty can only

come into play insofar as they are reflected in maxims—i.e., "normalized" in Herman's terms.

The strength of this model—its internal rational coherence—is also, in a sense, its weakness. It depends on accepting a claim that what most matters is moral reasoning and that the very structure of that reasoning itself places real restrictions on what can count as a reason, and so ultimately determines both what is permissible and required. The advantages of taking such a position are manifest—if it could be brought off it would prescribe moral activity on pain of irrationality. Thus moral requirements would acquire something like the force of the conclusions of valid deductive proofs—they would be escapable only to the degree that one was willing to give up being rational. The disadvantage, of course, is that the whole project depends on a series of claims about metaphysics of morals that are nothing if not controversial.

Faced with this, many philosophers have attempted to change the emphasis in Kant's moral thinking away from metaphysical claims about autonomy towards his more determinate conclusions. Their attention is thus diverted from the formula of autonomy, "Always choose in such a way that in the same volition the maxims of the choice are at the same time present as universal law" (440), towards the formula of humanity, "Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means" (429). We have to remember, of course, that Kant thinks these formulations are logically interchangeable, that each entails the other. Nonetheless, highlighting the formula of humanity, the emphasis is changed from how one structures one's willing to how one treats others. This gives rise to the Kantian

notion of "respect" for (the autonomy of) others, which has been perhaps his most important and lasting contribution to moral philosophy.

Love and Respect

I want now to look at how coming at Kant from this perspective affects our consideration of the demands of love and loyalty. In doing this I want to consider a bit of Allen Wood's book, *Kant's Ethical Thought*. In discussing Kant's thinking about love and friendship, Wood writes:

> Kant thinks there is a basic tension in human nature between loving people and respecting them… Respect and love are not mutually exclusive: In rational philanthropy they even go necessarily together. But in our natural inclinations they make for an unstable combination. Love as inclination is based on the pleasure we take in another (or her perfections), but respect for others pains us by striking down our self-conceit. Love is an empirical inclination, however, it is not opposed to self-conceit and indirectly it is even an expression of it. (Wood [1999], 271)

It is this tension that I want to explore.

The formula of humanity is supposed to prescribe the nature of our interaction with all other rational (i.e., moral) creatures. The foundation of that interaction is always the respect that one is required, on pain of irrationality, to pay to the autonomy of others. Thus while it is certainly possible to take a variety of affective attitudes towards others, such as being attracted or disgusted, etc., morality requires that they always be treated with the respect due to moral agents, i.e., to rational ends-setters. Ultimately, what we most fundamentally owe another is rational respect, regardless of what we might feel towards him. Thus, any feelings we might have towards him, positive or negative, cannot justify any abrogation of what respect requires.

This is an important point because this requirement prescribes the moral core of interpersonal interaction. No other considerations can ever impact or detract from the requirement of rational respect. In all regards then, my conduct towards my fellow man is to be guided by this requirement, and no matter whether I love or hate him I must respect him.

As Wood notes, this requirement is very much in tension with the apparent requirements of love. Respect requires a kind of distance; it forces us to cede to others the space in which to choose their own ends. Wood writes, "Perfect friendship requires mutual respect, which will be threatened by too much love, too much familiarity, too much openness" (282). Love is incompatible with respect precisely because it tends toward communion, the union of the lover and the beloved. Love breaks down the barriers between independent selves, causing us to lose sight of others' independence. As such, it threatens to overcome the distance necessary to recognize another's authorship of his own ends.

This danger seems to arise from the notion that love "is based on the pleasure we take in another (or her perfections)" and so "it is not opposed to self-conceit and indirectly it is even an expression of it." The idea here seems to be that because love is based on pleasure, to the extent that we give reign to our inclination to love we are merely gratifying ourselves—hence, expressing self-conceit. On the other hand, "respect for others pains us by striking down our self-conceit," presumably by making clear to us that other rational creatures can (and do) set ends very different from our own. Respect implicitly acknowledges others as independent sources of value thereby diminishing the value that we instinctively ascribe to ourselves.

To fully understand the reason Kant takes such a strained view of love we need to be aware of a distinction he draws concerning types of love. For Kant, 'love' can either refer to "love as an inclination" or "benevolence from duty," and it is only the latter kind of love which is "practical, not pathological" (399). Not only is the latter kind of "love" permissible, it is also a duty. On the other hand, Kant understands "love as an inclination" in terms of pleasure.[24] I suspect that many will find this dichotomy unsatisfying. I doubt many people, philosophers or otherwise, would count "benevolence from duty" a species of love at all, but at the same time most would also think that love has to do with something more than merely "pleasure in the reality of the object." Before I take this up, however, it will be useful to understand precisely what Kant thinks the problem with love "as an inclination" or pleasure is.

Kant calls "love as an inclination" not merely a temptation, but "pathological." The pathology lies in what acting on an inclination is supposed to do to the will.[25] When the will is determined to act from desire it becomes "heteronomous," such that it "does not give itself the law, but [its] object does so because of its relation to the will" (441). In other words, when motivated by a desire for pleasure the will is not free. It is being determined by something outside of itself—viz., the object of the desire.

---

[24] Allen Wood writes that in Kant "Love as inclination is based on the pleasure we take in another (or her perfections),… Love is an empirical inclination" (Wood [1999], 271). Wood refers us to the *Second Critique* where Kant writes that in desire "the determining ground of choice consists in the conception of an object and its relation to the subject…. Such a relation to the subject is called pleasure in the reality of the object…" (21).

[25] It is quite possible that Kant is using 'pathological' in its literal sense as well—i.e., as related to *pathos* or feeling. Nonetheless, since a will determined by pathos is morally defective for Kant, I think the other sense of 'pathological' is also in play here. I am indebted to Andrew Chignell for pointing out the sense related to pathos.

We must always remember that for Kant the deepest ground of morality is the freedom of the will. But the will is only free insofar as it is "a causality in accordance with immutable laws," as a "law to itself." Because the pure will is fully rational, it necessarily conceives of itself as law governed, and these laws must be a priori, both universal and necessary. In the *Second Critique*, Kant writes that when a will is determined by a desire for pleasure it "must be always empirical" because "we cannot know, a priori, from the idea of any object, whatever the nature of this idea, whether it will be associated with pleasure or displeasure or will be merely indifferent" (21). Pleasure cannot inform a universal and necessary law because the "subjective susceptibility to a pleasure or displeasure" can "never [be] known except empirically and cannot be valid in the same form for all rational beings." Thus, it "lacks objective necessity, which must be known a priori", and "such a principle can never furnish a practical law" (22).

The idea is relatively straightforward—no given object will necessarily elicit the same pleasure response in all rational creatures. Whether or not a given person takes pleasure in any given object is contingent and can only be discovered empirically, by exposing that agent to it. Accordingly, pleasure cannot determine a free will because a will is free only insofar as it is determined by necessary and universal laws which it gives to itself. Conversely, any will determined by pleasure is unfree, and since freedom is the ultimate ground of morality, a will determined by pleasure is, at least in part, "pathological" and immoral.

My chief concern here is that we understand that Kant's moral outlook commits him to the proposition that respect is the most fundamental attitude we must take to others. Thus, to take an attitude towards another based on

love or friendship alone is to disrespect him. Further, what I *owe* another, as opposed to am inclined to give, is ineluctably a consequence of respect alone. For I cannot act morally towards another, except insofar as I recognize in him his "humanity," i.e., as a cause of moral action, as a rational end-setter.

I think this way at looking at love, and interpersonal relations generally, is mistaken. When I act out of love, to the extent that it is a real love, it does not seem accurate at all to say that I am acting for "the pleasure [I] take in another (or her perfections)." This is a wildly unrealistic, reductionist account of love, equating it with pleasure. Let us go back to the example of the wife trapped in the burning building.

When the loving husband rushes into the building to save his wife he does this because he loves her. In other words, as an explanation of his action this fact—that he loves this woman—is a complete explanation. He does not do this because of the pleasure he feels in loving her, or in the prospective pain he would feel if she were lost. To try to reduce what is at stake in his will to a question of pleasure or pain seems to completely obscure what is at stake. The entire point of Williams' claim that any sort of calculation would be "one thought too many" is that the husband's will is directly determined to this action by his caring for his wife.

This does not mean, of course, that love puts reason in complete abeyance (though it may appear that way for some lovers), for the husband must still make rational plans and choices involving the best way to achieve his ends, etc. Rather, the point is that his end—to save his wife—is in some way determined for him by his love for her. There is simply no deliberation (i.e., practical reasoning) involved in this "choice" of end, much less any kind of consideration of what respect, either for his wife or for the stranger, would

require. Such a calculation of the demands of respect would reveal an agent who really did not love his wife to the degree that Williams imagines. What we need to see, however, is that what moves him is not the pleasure he takes in her (or her perfections); it is simply his love for her.

This will become clearer in Part II when we explore Frankfurt's very fruitful thinking about love. What is at stake here, and what we need to understand, is that there is a genuine tension between love and Kant's notion of respect.[26] However, as I will argue below, it is by no means the case, as Kant supposes, that the good will is necessarily the one in which respect always wins out over love. The tension is at heart a reflection of Kant's tendency to collapse any distinction between a perfectly good will and pure practical reason.

Respect is the proper and inescapable end of considering another only in view of her rational powers. However, when I love a woman, I do not love her rational powers (though, speaking *loosely*, I might say that is one of the things I love about her), nor do I merely revel in my own pleasure. I love *her*, such that her good becomes for me like my own—because indeed it becomes a part of my own. To say this, however, is not merely to expand my self-conceit as Wood describes it—what is aimed at is still *her* good[27]—even as her good becomes inescapably linked with my conception of my own.

Accordingly, when the husband rushes into the burning house to save his beloved wife, it is little different, with respect to his will, than if he were to

---

[26] This is a consequence of the structure of *Kant's* conception of respect and may not necessarily be intrinsic to a broader, different conception. I will take up such an alternative understanding of respect below.

[27] Of course, it is my conception of her good, and I could be mistaken. Nevertheless, in a case like this one, a hasty exit from the burning house is clearly good for her, and thus it is transparently her good at which the husband's action is directed.

do something dangerous or strenuous to save his own life. I cannot emphasize enough, however, that he is not acting as he does because of the pleasure he takes in her (though he might genuinely and legitimately take much pleasure in and from her), but because his will is determined towards her good by the very structure of his caring for her. Accordingly, to make his attitude towards her essentially one of respect is precisely to fail to express his own deepest convictions. It is, again, one thought too many.

The conflict then is about the fundamental attitude it is appropriate to take towards those to whom one has special relationships. Following Wood's Kant we would be led to conclude that towards everyone—loved one, stranger, or foe—our most fundamental attitude must be respect. On top of that respect, but never violating it, we might be thought to add layers of sentiment, emotional attraction or repulsion. However, Kant's picture of morality simply requires that any duties we have towards others proceed from the motive of respect alone.

While I am inclined to agree that we owe a certain kind of respect to all of our fellow humans, indeed to all of our fellow creatures, nonetheless I feel that taking, as most fundamental, the attitude of respect will oftentimes be horribly inappropriate. My most fundamental attitude towards my mother or my wife should not be one of respect; it should be an expression of a love so deep that it is constitutive of my very will. This does not mean that I cannot or indeed should not respect the "humanity" of my mother or my wife, but it does seem as if that respect is oftentimes a secondary consideration. Of these sorts of objects of my love, classing them together with the mass of humanity seems to radically undervalue them. It is almost a category mistake.

A Suggestion: Re-construing the Relationship of Respect to Love

I fear that some readers may think I have rather undervalued the importance of respect. I want to be clear that the deep incompatibility I have been discussing thus far is a consequence of the nature of Kant's understanding of respect. We must continually bear in mind that in Kant's system the demand for respect is integral to his account of autonomy; they are, in a sense, the flip sides of the same coin. As I noted above, the demand for respect is intrinsic to rational agency in Kant's system.

The deep conflict might be resolvable to the extent that we offer a broader understanding of autonomy. If autonomy encompasses more than merely the agent as an instantiation of rational agency, we might be able to construct an account of the demand for respect that can integrate our intuitions about the (moral) importance of love. In the next chapter I am going to examine Frankfurt's account of autonomy at length, so it will not be my purpose to do so here. Nonetheless, as a kind of preview I want to offer a suggestion (and that is all it will be at this point) of how autonomy and love, and thus respect and love, can be reconciled by rejecting Kant's idea that agency is grounded in practical reason alone.

As we have seen, for Kant the will can only be free insofar as it gives itself necessary laws; that is, insofar as it is determined by immutable laws that hold for all other rational wills. For Kant the will is essentially (and exclusively) rational and that rationality can only be fully expressed in a priori laws. Love is incompatible with autonomy precisely because it is "empirical"; it cannot underwrite necessary and universal laws. The question we have to ask now is simple: is Kant's understanding of autonomy correct?

Let us consider Harry Frankfurt's alternative understanding of autonomy. Rather than understanding the will to be essentially its rational nature, Frankfurt has a theory of the structure of the will as divided into a hierarchy of volitions or desires. First-order desires are desires for particular objects external to the will (e.g., the desire for food). Second-order desires are desires whose objects are first-order desires. A paradigm example: the desire for a cigarette is a first-order desire; the desire to not desire cigarettes (e.g., as part of a commitment to stop smoking) is a second-order desire. Frankfurt holds that autonomy consists in forming effective volitions according to one's second-order desires.

Frankfurt writes:

> …the statement that a person enjoys freedom of the will means (… roughly) that he is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or have the will he wants…
>
> It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will… (Frankfurt [1971], 20).

For Frankfurt, autonomy consists in having the effective will that one wants to have. It does not require that I want a particular kind of will (e.g., one that desires that it be determined by necessary and universal laws); it only requires that I be committed to having the will I do, in fact, have. There is an important proviso; the autonomous agent needs to be reflectively committed to his second-order volitions. For it is those second-order volitions which will determine which of his first-order volitions with which he "identifies himself *decisively*" (21).

A central idea of Frankfurt's conception is that it is this decisive commitment to *identify* with some first-order volitions and to reject others that is at the heart of free agency. This notion of identification is not accidental. Who the agent is, what he is, is determined by what he identifies himself with. The process of identification, a commitment that "resounds" through his will, is at the core of his autonomy, because it is precisely the having the will he wants to have.

Under Frankfurt's account of autonomy, there is no problem reconciling love and autonomy. On that account a will determined by the love for some object is free just insofar as that determination is the one the agent wants to have. In other words, a will determined by love is autonomous just in case the agent wants to love what he does in fact love, if his love is expressive of his second-order volitions.

We should note that this result is neither tautological nor trivial. It is possible, I take it, to have loves that one does not want to have. Someone, for instance, who grew up in an abusive home might still love his abusive parent (he might still have some complicated set of first-order desires towards the parent), even though he would say, on reflection, that he does not want to have such a love. Similarly, a jilted lover might wish that he no longer loved the woman who had left him, and yet I think we would be inclined to say that, at least in certain circumstances, he still loves her. In a third situation, someone might love something, such as kitsch, that he thinks is unworthy of love and wants not to love. In such cases, the agent is not free with respect to his loving, and accordingly if he were to act on these first-order volitions he would be acting with a divided will. He would be unfree in a manner analogous to what Frankfurt calls the "unwilling addict"—a person who acts

to gratify his addiction while simultaneously wanting not to have the first-order desires which constitute that addiction.

To freely love, then, is to identify with the desires that constitute that love. If that is the case, then somehow it is in my loving that I express who I truly am. On Frankfurt's account the will itself is partially constituted by its "core commitments," i.e., what it most cares about. For it is these core commitments, these higher-order volitions, which regulate and determine the agent's free actions. The agent expresses his freedom in acting upon these core commitments, and so they can truly be said to constitute, at least partially, his identity and the ground of his free agency. Thus, loving is not only compatible with autonomy, but can itself be an expression of it.

Now I want to note some of the ways in which this alternate account of autonomy seems superior. First, I think it is important to see how this account of the will lets us link love to the will at the deepest level. Thus it allows us to see love not as a mere contingent passion, but as issuing from the identity of the agent himself. I think this does the most justice to the phenomenology of deep loves, where the good of the object of his love seems to become so integrated in the agent's self that an insult to the object is an insult to the agent. In other words, since freely loving X consists in desiring the good of X and wanting to desire the good of X, for something bad to happen to X is for one of my own desires, especially one I identify with, to be frustrated.

Beyond how it feels, however, as a theory of the will this view has a significant advantage in how it construes love. Under a Kantian view love necessarily takes on a secondary role. On such view love does not—cannot—determine the will freely (i.e., a will determined by love can never be fully

autonomous). Thus, on such a view my love is less about me, than about the object of my love. Yet in the deepest loves it seems as if my love issues from the very core of my being. It is at this point that we begin to descend (or ascend depending on how one looks at the matter) into the realm of poetry. But, I take it that the consistency of the metaphor that poets have used to express this point suggests something very important about the phenomena itself and to which the Kantian account seems woefully inadequate.

Kant thinks that the pure will is essentially rational, thus it cannot be determined by anything less than a necessary and universal law and remain free. However, I do not think we have to follow Kant here. Although it is certainly the case that our rationality is an important part of our volitional structure, it does not seem to me correct to assume that my rationality alone is who or what I essentially am. In fact, insofar as my will is merely rational it must as well be in a sense anonymous, for whatever laws it gives to itself apply to the same degree to all other rational wills. There is nothing about my will that makes it mine besides the bare "I think" that Kant notes, in another context, attach to our cognitions.[28]

Delving into the metaphysics of identity is obviously far beyond the scope of the present effort. However, I do want to suggest that the relationship of love to identity matters and supports the type of view I have sketched above. In short, if love can be an expression of a free will, then we can love another person without necessarily failing to respect her. On the account I have sketched, I am not limited merely to respecting her as another instantiation of rationality, but I am free to love her for what makes her who

[28] Williams discusses this idea of the anonymity of the Kantian agent in chapter 4 of *Ethics and the Limits of Philosophy*, (Williams [1985], 68-70).

she freely is—namely, the sorts of things that she loves, which make up her own core commitments.

None of this, we should note, implies that either the rationality of the agent or that of his beloved is unimportant. A will constituted by commitments that are rationally incompatible is incoherent, and quite possibly wantonly unfree. We can and do subject our core commitments to rational reflection, and it is quite proper that we do so. On this view of love and autonomy rationality is important, it just is not alone the ground of all moral agency. The composite will, regulated by the requirements of rationality, but also expressing its internal structure as constituted by its higher-order volitions, is the true ground of agency.

This then brings us full circle. We began by wondering if love was compatible with respect. I suggested before that our concern with respect for persons actually arises out of Kant's concern for securing the autonomy of rational agents. Respect, for Kant, is a logical consequence of rational autonomy.

The account I have offered reconciles respect and love, basically by rejecting Kant's view of autonomy. It is important to note that I have not rejected the importance of autonomy, nor its position as the ground of morality (I can actually remain agnostic on the truth of the latter proposition). Rather, I have tried to suggest that love and respect can be brought together if we realize that respect does not require us to recognize in another merely her pure capacity for rationality.

In fact, the view offered here suggests that respect consists in seeing another for what he actually is, what is expressive of his will. If Frankfurt is correct, while his rationality will be a central part of an agent's will, it will not

constitute it entirely, nor will it necessarily be the most important thing about that will. Respect for another will require us to see the agent for what he is, and this will not be merely a formal property but will demand that we actually know something substantive about him, namely his loves and (especially his higher-order) desires. Obviously, much more can be said about the nature of respect informed by this notion of autonomy, but we must close for now.[29]

My "suggestion" has taken us rather far afield. My purpose here was less to discuss the interesting interconnections of autonomy, respect, and love fully and more to offer the sketch of an alternative account that contrasts starkly with the Kantian picture. If we thought that a Kantian account of autonomy were the only one on offer, then we might be more willing to follow him in devaluing love. What I hope to have accomplished in this discursion is not a complete solution of the tension between love and respect, but to have strongly suggested that one is possible if we give up Kant's assumption that only a will determined by a respect for the moral law alone can be fully free.

The issues surrounding the proper understanding of *respect* go well beyond Kant exegesis proper. As I noted above, *respect* has become a central consideration in moral philosophy generally in, for example, contemporary liberal thinking following Rawls. Therefore, I want now to turn to one of the most important outgrowths of the Kantian project which is presently engaging moral philosophers—moral contractualism.

---

[29] Frankfurt's discussions in "Equality as a Moral Ideal", "Equality and Respect", as well as Raz's remarks in reply ("On Frankfurt's Explanation of Respect for People") show something about how such an account might go, but much more remains to be done in exploring this terrain.

Moral Contractualism

The project of moral contractualism seems to have developed, broadly, out of some of the issues Rawls has brought to the forefront of moral philosophy in the last 30 years. Other philosophers have taken up and expanded his original social and political project into the realm of normative ethics and metaethics. One of the more important of these philosophers is T. M. Scanlon, whose recent book *What We Owe To Each Other*, serves to collect and develop various strands in the developing project of moral contractualism, and whose work will serve as springboard for my discussion here.

Comparing his own work to Kant's, Scanlon notes that while "Kant sought to explain the special authority of moral requirements by showing how they are grounded in the conditions of our rational agency, I try to explain the distinctive importance and authority of the requirements of justifiability to others by showing how other aspects of our lives and our relations with others involve this idea" (Scanlon [1998], 6). Accordingly, Scanlon acknowledges that his view "is, in Kant's terms, avowedly heteronomous."

Obviously, my goal cannot be a comprehensive examination of moral contractualism, within which there are now various live disagreements, or even just of Scanlon's own detailed work. Rather, my comments here will be directed at what I take to be some of the most fundamental premises of the moral contractualist project, again asking the same kind of over-riding question we have considered vis-à-vis consequentialism and Kant: what can moral contractualism say about the demands of love and loyalty?

The core claim in the project of moral contractualism concerns the justification of moral principles that are binding on a group of agents. That claim is simply that judgments of right and wrong "are judgments about what would be permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of the behavior that others, similarly motivated, could not reasonably reject" (4). Thus, what centrally characterizes the moral contractualist project is an attempt to find principles which are justifiable to all the parties involved. The essence of this test of justification is embodied in a requirement for those who oppose a proposed principle to provide reasons for their objection, which everyone else can accept *as a reason*.

Moral contractualism draws on several strong intuitions, and its plausibility is directly proportional to the degree in which one is committed to those intuitions. These importantly include a claim about the primacy of reasons, and a thoroughgoing individualism. What ultimately matters according to this view is that representative (generic) individuals agree to principles binding them all, by accepting any principle against which there are no objectively compelling reasons. In other words, for a principle to be objectionable it must offend against the interests of a *generic* individual in such a way as to generate reasons against it that are universally available to all other agents.[30]

[30] Scanlon clarifies his position on page 171: "To avoid gridlock we must move away from the idea that each person's life or each person's happiness 'matters' to the question of an acceptable system of general principles of action. Acceptable principles could not require us, in deciding what to do, to consider how every actual individual would feel about it. And in deciding which systems of principles are 'acceptable,' we cannot envisage the reactions of every actual person, We can consider only representative cases, and take into account only those objections that a person could raise while recognizing the force of similar objections by others."

Just as it is the case with social contract theories that no claim is made about actual constitutional conventions, so with moral contractualism there is no claim about the interests of actual specific individuals, variously motivated. On the contrary, what counts in this view are those reasons which are available to everyone, such that for a principle to count as objectionable, it must be so in virtue of reasons that an individual not in the position of one of those most directly affected can appreciate.

On this view, as Scanlon says, "the idea of a reason [is] primitive" (17). This is very important because the constraints on what can count as a reason ultimately become constraints on the limits of morality. The central constraint, built into the core commitment, is that a reason cannot make reference only to the particularities of a given *real* individual's desires, but must be available to all as a consideration shared by all *generic* individuals of a given class.

I want now to return to our motivating question: what can moral contractualism say about the demands of love and loyalty? One thing we should note, off the bat, is that Scanlon is not committed to asserting that moral contractualism picks out all of what we refer to with the word 'morality.' There may be loose uses of the word that attempt to pick out other situations in which behavior is guided by principles possessing some kind of authority that seems similar to the authority of morality, but Scanlon can claim that such uses actually operate under values outside his specific concern with rightness and wrongness. This being the case, there may be genuine values which the moral contractualist constraint on justifiability does not precisely pick out. Nonetheless, whatever these other values are, they cannot, in themselves, override the requirements of moral contractualism. In other

words, while there might be actions guided by genuine values either outside or on the periphery of morality proper, insofar as any of those values find expression in activities that contravene the requirements of the justifiability criterion, such activities are illegitimate and must give way before the requirements of morality as articulated by the justifiability criterion.

Scanlon considers these issues under what he calls the "problem of priority," which is "the question of how the morality of right and wrong is related to our other values and how it could make sense to give it priority over them" (160). Scanlon chooses to examine friendship, and how the particular demands of friendship (e.g., to favor this particular person over a similarly qualified stranger in the dispensation of some benefit) can be reconciled with the priority of "the morality of right and wrong."

Scanlon asks us to consider the following example: suppose that one's kidneys have failed, and in an attempt to save your life one of your friends attacks a stranger and steals his kidneys to replace yours. Such a friend, claims Scanlon, would not be what we would be willing to call a friend at all, because friendship "involves recognizing the friend as a separate person with moral standing" such that a "person who only saw friends as having [moral] status" merely in virtue of being a friend, "would not have friends in [Scanlon's sense]: their moral standing would be too dependent on the contingent fact of his affection" (164).

This example is supposed to show "that friendship… requires us to recognize our friends as having moral standing as persons, independent of our friendship, which also places limits on our behavior" (165). In other words, "the conception of friendship that we understand and have reason to value involves recognizing the moral claims of friends *qua* persons, hence the

moral claims of nonfriends as well… [such that] no sacrifice of friendship is involved when I refuse to violate the rights of strangers in order to help my friend. Compatibility with the demands of interpersonal morality is built into the value of friendship itself" (*ibid.*).

What we need to take out of this discussion is that idea that the *moral* claims of friends are only moral insofar as those claims are made *qua* persons. In other words, according to Scanlon, necessarily contained within the kind of friendship we "have reason to value" is the background assumption of the person-hood of all the parties involved, friend and non-friend alike. The requirements of morality concern their status as persons, not as friends, and so can only bind in virtue of that status. It should be clear, of course, that being a person just means being "someone to whom justification is owned in his or her own right."

We should instantly recognize this requirement as the same requirement made by Kant's formula of humanity—namely, the demand for respect. The idea underlying this is that morality inescapably requires respect for all persons, irrespective of whatever feelings one bears them, positive or negative. The priority claim in Scanlon is supposed to reconcile the apparent conflicts between the demands of narrow morality and the demands of love and loyalty by somehow showing that the demands of love and loyalty, as in the demands of friendship in the example at hand, presuppose, or have built-in to them, respect for persons.[31]

Scanlon is very clear-eyed in understanding this feature to be necessarily characteristic of "friendship as we understand it, or with the

[31] "I believe that what I have argued here in the case of friendship is true as well of other personal relations whose demand many seem to conflict with morality, such as family ties and relation with other members of a team or cooperative enterprise" (166).

conception that we have most reason to value" (165). In other words, there are other possible conceptions of friendship (Scanlon cites Achilles and Patroclus), which would not necessarily have this feature; but, claims Scanlon, those other conceptions would either be alien to us, or we would not "have most reason to value" them. This presumably means that the only demands of love and loyalty that we have reason to value would have to have this feature (respect for persons), because of the nature of our more fundamental commitment to justifiability. This leads Scanlon to embrace a kind of relativism, as acknowledges that "the degree to which there is a conflict between the morality of right and wrong and the goods of personal relations depends greatly on the society in which one lives" (166). At the same time, however, he claims that we *do* have and value the sort of friendship he describes (where respect for persons is "built-in") so that, *for us* (i.e., for those persons having this conception of friendship) the matter will always be resolved in favor of the priority of morality.[32]

It should be no surprise that I will take issue with Scanlon's conception of friendship in the same way that I took issue with Kant's conception of love—namely, by doubting whether a respect for persons really is the most fundamental issue in the relationships which give rise to conclusive demands

[32] When I speak of *relativism* here, we must acknowledge that Scanlon is not necessarily committed to the notion that the "morality of right and wrong" is relativistic, but rather that the scope of the conflict between that "morality" and the "goods of personal relations" will be relative with respect to one's society. It is not entirely clear how far down this relativism would go. In other words, what would be the status of the demand for justifiability in a society in which nobody had a pre-existing commitment to it, or for that matter, to the giving of reasons at all? I want to be clear, however, that I do not disagree with the truism that at their very deepest level, to some degree, the nature of our commitments is relative to the society in which we live. My main disagreement with Scanlon on this point is whether we live in a society in which the conception of friendship he has sketched (centrally involving a respect of persons) really does accurately capture what characterizes real friendship in our society.

of love and loyalty. As I noted above, there is a deep tension between the demands of love and the demand for respect. The requirement to see our loved ones, first as persons, and only then as the objects of our love, presupposes the primacy of our commitment to morality in a way I believe inconsistent with an ordering of the will that gives adequate space to the demands of love and loyalty.

Scanlon himself specifically mentions Williams' example of the husband choosing to save his wife from the burning building, but he seems rather unconcerned about it. He notes that the example merely gives rise to further questions about the priority of the demands of love and loyalty and their relationship to morality proper. As we have already seen, although he is willing to acknowledge that the demands of love and loyalty proceed from values that are worthwhile in themselves, Scanlon also holds that insofar as the demands of love and loyalty conflict with what morality requires, the former must give way. The reason for this is that the demands of morality—specifically, the requirement of justification—are more basic than the demands of love and loyalty, because respect for persons is "built-in" to all our interpersonal relations, such that the demands of respect must trump all other demands when they conflict.

Despite noting Williams' example, I do not think Scanlon has properly appreciated its force. As I have tried to show above, the point of the "one thought too many" complaint is that justification is simply out of place in this context. The idea that Williams' (and later Frankfurt, who embraces the example)[33] is trying to get across is that for the husband there simply cannot

[33] See Frankfurt [2001], "Some Mysteries of Love."

be any deliberation about ends in this situation. His will is directly determined by his love, in such a way that he simply must go after her. To demand justification for this action is simply to misconstrue its nature, because the movement of his will simply has very little, if anything, to do with morality.

The moral contractualist might respond that this mistakes his project, because what is at stake is not the justification of any given *action* to others, but rather a justification of the *principles* guiding action. Thus, the moral contractualist will still maintain that any genuine value in the demands of love and loyalty can be accommodated in principles that everyone can accept, because recognizing the value of having loving relationships does not require one to actually refer to particular objects. Thus, generic individuals would be able to recognize the value of having loving relationships in general, and thus this fact could become a reason for rejecting principles that would undermine such relationships—all of this without making reference to particular objects (i.e., particular beloveds).

I think this still misses the point, however. Because what is at stake is leaving open the possibility that an individual's will could be determined in such a way that his commitment to morality would not necessarily be always overriding. This is of course connected with my claim above, that the valuing involved in the determination of the will that the husband experiences towards his beloved wife does not depend on her person-hood. I think Scanlon is simply wrong in claiming that respect for persons is built-in to all of the conceptions of special relationships we have reason to value, precisely because this fails to takes seriously the volitional requirements of love.

A fuller clarification of these issues must await Part II when we examine the nature of the loving will directly; however, we can preview what

is at stake by considering another answer to Scanlon's priority question. Scanlon evidently thinks our commitment to morality means that respect for persons is basic to all of our conceptions of interpersonal relationships that "we have most reason to value." Is there an alternative view?

Indeed there is; Frankfurt's view of the place of morality seems to me to capture more closely what is at stake here. In a response to Barbara Herman's complaint that he cannot account for the objectivity of moral values, Frankfurt claims that "we act morally when we are moved by love for a certain kind of world or a certain kind of life" (Buss and Overton [2002], 277). What we need to recognize here is the nature of Frankfurt's novel claim. Frankfurt is basically saying that there is a more fundamental structural consideration than the requirements of morality, which is of course the structure of the will. The novelty of the claim, and I believe its deep truth, lies in the notion that our moral activity proceeds from a more basic volitional commitment to a particular kind of world or life. In other words, what is basic here is not respect for persons, but a particular structure of the will.

As I will be at some pains to argue in Part II, it seems to me that the demands of love and loyalty, properly understood, are also determinations of the will. Our commitments to what we love are constitutive of the will in such a way that we are determined by them with respect to certain actions. Now this does not mean that the demands of love and loyalty will always over-ride the demands of morality, because to the extent that we are moral people we are also committed to (i.e., in love with) morality. Thus, the conflict that plays out will not, as Scanlon asserts, necessarily play itself out in the space determined by the requirements of morality (as described by his system of justifiability or any other), but in the space whose shape is determined by

what the given agent loves. The conflict will be analogous to the conflict in the will that occurs when two objects of my love are pulling me in different directions.

The situation, then, is more like which obtains when the two people in the burning house are my wife and my daughter (instead of the stranger), both of whom I love unconditionally. In that case, the choice between them becomes almost debilitating, precisely because there is no sufficient reason to choose one over the other, and indeed could not be. What is at stake is a kind of schism in the will itself, where the nature of my loving commitments each demand unconditional obedience, but which both cannot be accommodated. The conflict then, is a conflict of loves, and if there is a resolution it is because one love wins out over the other.[34]

Thus when the demands of love and loyalty conflict with the demands of morality what is at stake is a conflict of loves in the same way—my love for this particular person (or institution, etc.) versus my love for a certain kind of world or way of life. In a moral person sometimes, perhaps even the vast majority of the time, his love for morality will triumph over the love of a particular beloved; but, sometimes, it may come to pass that the love of the particular will override a consideration of the requirements of morality. In other words, Scanlon is mistaken in thinking that respect for persons is built-in to all the conceptions of interpersonal relationships that we do actually

[34] It seems to me that in this situation the state of the will is very complicated. This could result in no action being taken at all, because my will, in the face of these incompatible requirements, simply breaks down, so to speak. Additionally, I might, rationally, conclude that the choice situation is ambiguous and then go after one versus the other for no good reason at all, but this also represents a kind of break down. The particulars of the case, for the particular husband and father involved, will be determinate.

value.[35] Indeed, the proper expression of the value of those relationships will often require the they conflict with the requirements of morality.

I think a kind of rescue case can make this clearer. Consider the stock example of being in a position to rescue the inhabitants of only one of two islands, when both are threatened (e.g., by some natural disaster) and it is impossible to get to both in time. If there are an equal number of persons on both islands, then whichever group that is not saved is generally not taken to have a valid moral complaint against the rescuer. Thus if there is one person on each island, if the rescuer flips a coin and rescues the person on island A, leaving the inhabitant of island B to perish, it is generally conceded that the person on B does not have a moral complaint against the rescuer.

On the other hand, if island A has two inhabitants and island B only one, then it would seem that equal respect for persons would require the rescuer to go to island A. The reason for this is that if he goes to island B, he will seem to have wronged one person.[36] Thus, it would seem the agent would be wrong to, for example, flip a coin to determine his action, because the greater number of persons on island A provide a sufficient reason for rescuing them, while giving equal respect to everyone.

Consider the case if the single person on island B is the rescuer's beloved mother. Seeing her danger he feels volitionally compelled to save her and does so. In such a case, he has not treated all the victims with equal

---

[35] As to whether they are built into all the conceptions we "have *reason* to value," that is a sufficiently difficult question that I will leave it here unaddressed. In any case, I am content with the claim that the actual and real demands of love and loyalty towards which our intuitions aim (as brought out in Williams' example) are embodied in the values we actually *do* place on our beloveds.

[36] Only one and not two persons are wronged because one of the two is "cancelled out" in some way by the person on island B who was rescued. The example presupposes generic individuals, and no pre-existing claim by any of them on the rescuer.

respect. At the same time, however, I think many observers would not consider the rescuer's action to be wrong, certainly not in the same way had he gone to island B on account of a coin flip. Many of us would not only understand the rescuer's action, but consider him to have done nothing blameworthy. I concede that others would think him wrong, of course, but I consider the example plausible enough, even for those who disagree, to show that we can and do have values attaching to particulars which are deeper than equal respect for persons.[37]

Bear in mind that I am not suggesting that the fact that the person on island B is the rescuer's mother gives him a better *reason* for going to B. Rather, I am suggesting that if his love for his mother wins out over his love for an impartial morality he is far from being a moral monster. While others might disagree, I cannot help but think that many would not only *not* blame him for rescuing his mother but would consider him in some way inhuman if calmly disregards her for the sake of cool impartial reason. The point is that the agent's available reasons are not alone the ultimate deciding factor in evaluating good action.[38]

---

[37] I would not be surprised at all if intuitions would change where the numbers different. If, for instance, there were a thousand people on island A and still only the rescuer's mother on island B, then we probably would be less inclined to view his saving only her indulgently. I do not wish to suggest that there is some precise equivalence where one's deep attachment to one's mother outweighs the claims on one other person, but not 999. It is sufficient for my point here if the reader at least takes it as plausible that in the two versus one case presented the rescuer has acted reasonably. I am indebted to Richard Miller for suggesting these types of examples.

[38] As an aside, I think that the virtue ethics tradition could encompass this sort of example, insofar as the rescuer who goes after his mother could be considered a man of good character. I am not suggesting that he necessarily will be so considered by all virtue ethicists, but it is certainly an open possibility within a virtue ethics type view in a way it seems not to be an open possibility in a rationalist Kantian or moral contractualist type view. Nonetheless, it is only an aside and it is not my purpose here to make this argument for virtue ethics.

I want to be clear here, this is not a claim that the demands of love and loyalty *necessarily* take priority over the demands of the morality. Rather, this is a repudiation of the reverse claim, that morality *always* takes priority over the demands of love and loyalty. The point is that the authority of both of these sorts of demands *within the will*, insofar as they determine action, lies in the strength of our love for the values that inform them. In other words, love (in Frankfurt's sense of the constitutive commitments of the will), not morality, is what is basic. The conflicts between morality and the demands of love and loyalty play themselves out on a field whose contours are defined, not by morality alone, but by what we love (among which can and should be included considerations of impartial morality).

The point of the example is *not* that the rescuer is necessarily doing the right thing in going after his mother. Rather, this case is meant to counteract the intuition elicited by Scanlon's kidney-stealing friend example. There Scanlon claims that a friend willing to radically violate respect on account of friendship would not be the kind of friendship we have reason to value. The reason for that, according to Scanlon, is that respect is more basic than attachment within *our* concept of friendship.

What the mother example shows is that respect is not always more basic, that there are attachments which can inform an agent's actions without making that agent appear to be a moral monster. The power of Scanlon's example lies in how strongly it elicits our intuition that we cannot so greatly wrong a stranger on account of a friend. It does not show, however, that we cannot legitimately "wrong" (at least from an impartial standpoint) someone else on account of a deep attachment. As I suggested above, not only do I think it likely that many would not disapprove the rescuer going after his

mother, I think many would find positively distasteful his failing so rescue her. I will be the first to admit that this does not settle the moral issue at hand, but it does show that we have powerful intuitions suggesting that acting on impartial respect is not always best.

To claim, as the moral contractualist does, that morality is more basic, is to fail to appreciate the degree to which morality is dependent on a volitional commitment for its authority. Scanlon, in particular, should be sensitive to this because his claims so often make reference to the nature of *our* commitment to justification, with the full knowledge and recognition that people in other cultures and societies may have wholly incompatible commitments that do not include justification in terms of public reasons.[39] Accordingly, he should recognize how the demands of morality, insofar as they have authority—that is, insofar as they are real *demands*—are such only in virtue of our volitional commitment to them and to the ideals they embody. It is the commitments of the will (generally including a commitment to morality) that are basic, not the demands of morality *simpliciter*.

The plausibility of the foregoing, and indeed the plausibility of my argument for the legitimacy of the demands of love and loyalty as a whole, is wholly dependent on the plausibility of the picture of the will I am taking up from Frankfurt and Williams. Many of the considerations to which I have

[39] "If no one in my society understand friendship as having the moral content I have just described [i.e., respect for persons], then a relationship with others on this footing is not available to me. If everyone in my society sees the world as divided between 'them,' the outsiders to whom nothing is owed, and 'us,' who are bound by relations of blood, affection, and patronage, then I really am faced with a choice between actual ties with my fellow citizens—strong and warm, perhaps, if also fierce—and the requirements of morality, grounded in an ideal of relations with others that must remain purely ideal. I have tried to argue that we are not in fact faced with this choice, but it must be conceded that others could be." (166)

repeatedly appealed in these last two chapters only have force insofar as the claims I have made about the structure of the will and the nature of authority (as volitional) are sustained. It is now finally time to begin Part II in which I will turn to a detailed examination of that picture of the will, and the relationship within it between the demands of morality and the demands of love and loyalty.

Part II

CHAPTER V

Frankfurt and a Volitional Theory of Value

In Part I, I tried to show how and why the conventional universalist theories of morality are deeply insensitive to the demands of love and loyalty. I argued that the demands of love and loyalty are directed at particularist values for which those theories can not adequately account. In Part II, I will turn my attention to giving an analysis of those values aimed at articulating the nature of those values and their central place within moral psychology, both descriptively and normatively.

I have repeatedly appealed to Williams' example of the man acting to save only his wife, trapped in with a stranger in a burning building, without pausing to consider whether he was justified in doing so. I want now to remind the reader of Williams' conclusion which I quoted in the Introduction, as my discussion in this chapter will focus in large part on expanding on the very suggestive points he makes there. He writes:

> …But the point is that somewhere (and if not in this case, where?) one reaches the necessity that such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view, and that they also run the risk of offending against it.
>
> They run that risk if they exist at all; yet unless such things exist, there will not be enough substance or conviction in a man's life to compel his allegiance to life itself. Life has to have

> substance if anything is to have sense, including adherence to the impartial system; but if it has substance, then it cannot grant supreme importance to the impartial system, and that system's hold on it will be, at the limit, insecure.
> (Williams [1982], 18)

The main issue at stake in this chapter is simply: what is entailed by this requirement that a life have "substance"? Williams gives us this highly intriguing suggestion that an agent's commitment to impartial morality is structurally insecure because it is necessarily in conflict with other commitments which are intrinsically incompatible with it. Furthermore, these other commitments ground "allegiance to life itself." However, as suggestive as this is, he leaves us wondering how exactly these "deep attachments to other persons" (and, I would add, deep attachments to other non-personal objects like institutions, traditions, or ideals) play themselves out within a life of "substance."

In some sense, this entire chapter—indeed this entire dissertation—is an attempt to understand what a life with substance is, what values it embodies, and what possessing those values entails by way of action and belief. To begin to answer this question, however, I will leave Williams and turn instead to an analysis offered by Harry Frankfurt. It is in Frankfurt's rather radical work on the will, love, and especially what he calls "volitional necessity" that we will begin to find our way forward.

Frankfurt

Harry Frankfurt's work in moral psychology, ethics, and the philosophy of agency over the past 30 years has been nothing short of revolutionary. Perhaps most remarkable has been his rethinking of what

constitutes autonomy and personhood by developing an extremely influential hierarchical model of the will. In his more recent work he has moved even further into the confused thicket of moral psychology, and sought to (re-) introduce important categories of phenomena which have been overlooked or de-emphasized within the (overly) rationalistic tendencies of what he has called the "pan-moralism" of most modern moral theories.[1] These categories are essentially related to what we love and the way in which our wills, organized around our loving, are partially constituted by that loving.

I want to begin by examining some of Frankfurt's thinking on these topics, especially the nature of what he calls "volitional necessity." I hope to show how the deepest loyalties generate demands of love and loyalty which are a species of these volitional necessities and further how such volitional necessities are an inescapable consequence of the expression of the sorts of volitional values which inform the deepest loyalties.

---

[1] "It seems to me that many philosophers, in their approach to the task of identifying and understanding the norms that most authoritatively govern human conduct, are excessively preoccupied with *morality*. In their efforts to clarify and justify the basic standard to which we must submit when we act, they tend to focus their attention more or less exclusively upon issues pertaining to the requirements of moral obligation. Varieties of practical normativity that appear to be non-moral are either neglected entirely; or it is argued that in fact they are ultimately derivative from moral considerations, and that therefore their strictures concerning how we should act are just special cases of moral constraints. In my opinion, this pan-moralistic conception of practical normativity is mistaken. There are fundamentally authoritative practical norms that do not depend for their authority upon the requirements of morality. What I have particularly in mind are the requirements of *love*." (Frankfurt [1998], 4)

Volitional Necessity

This concept of volitional necessity was introduced by Frankfurt within an analysis of the relationship of freedom to necessity. Frankfurt was at pains to show how the contemporary "ideal of freedom," through which we are "fundamentally committed to encouraging a steady expansion of the range of options from which people can select," (Frankfurt [1999], 108) misses a vitally important point about the relationship between freedom and necessity. Namely, "as the ideal of freedom is more closely approached, the progressive reduction of necessity tends to undermine that ideal; and it also tends to undermine the ideal of individuality. For it is true both of freedom and individuality that they *require* necessity" (109).

Recall that Frankfurt has a hierarchical model of the will. An agent has first-order desire directed at objects in the world (e.g. hunger) as well as higher-order desires directed at lower-order desires (e.g. the desire not to be hungry). An agent is free, for Frankfurt, just in case his actions reflect his higher-order desires, i.e., what he most cares about. It is these higher-order desires or commitments which (partially) constitute the agent's will, because it is those commitments that determine who the agent most authentically is. Who I am is determined (in large part) by what I most love.

According to Frankfurt the dependency between freedom and necessity is structural. It is only in virtue of a certain sort of necessity that an individual can exercise the faculty of choice in a manner that represents an authentic expression of his own will. The "extensive growth in the variety of a person's options may weaken his sense of his identity." The reason for this is simple: "without a definitive set of goals, preferences, or other principles of choice" a

person's will itself will lack the kind of stability necessary for any given choice to be a genuine reflection of that will. His will itself will be in flux, and he "will be in a position to redesign his own will." However, once that happens no choice "will be fully or wholeheartedly his own" and as such he "is not volitionally equipped to make truly autonomous choices" (110).

The way in which this generates a necessity lies in the way in which such volitionally constitutive commitments determine the will. In other words, the agent's higher order commitments will express themselves precisely in binding the agent, in a sense forcing him onto a certain path, not as an external compulsion that limits his freedom, but as an internal impulsion that expresses it. I will examine this in more detail below, but to understand why this works this way we need first to understand the volitional structure that generates this necessity.

Without a stable core of higher order commitments by which he organizes his lower order desires, identifying with some and rejecting others, there can be no genuine *choices*. Of course, the individual may still act, after a certain fashion, pursuing the objects of various desires as they come to the forefront of his psychic attention. However, if this is the case there is no longer term plan or organization to his life; he becomes what Frankfurt earlier characterized as a "wanton," buffeted by the changing winds of desires, driven from one object to another without an enduring self-conscious sense of self. His unity is merely the unity of concatenation. What he lacks is narrative unity, the unity of a life responding to reflective commitments to principles and ideals that guide, organize, and so provide meaning to his life.

This stability is at the heart of genuine choice insofar that all such choices must be made in reference to some relatively fixed (in the context of a

given choice) identity determined by what one most cares about. Without such a fixity of his will, the agent's "choice" would fail to reflect anything enduringly particular about his will. Thus the modern ideal of a steadily expanding range of options encourages a substitution of whim over truly self-expressive choice. It is not that there is any kind of necessarily analytic relationship between an increase in the number of options and the decline of authentic agency. Rather the point is that the steady expansion of options tends to undermine the development of a self with genuine commitments (and thus a core identity) because it fosters a kind of flitting about, moving from one option to another and never acquiring the kind of understanding that can only come from long acquaintance and experience, of having to "live with" one's choices.

The key point that Frankfurt is after here is the idea that autonomy does not merely consist in the absence of outside compulsion, nor is it simply advanced by an expansion of options. Rather, autonomy consists in having a life organized around central volitional commitments, which themselves are not subject to easy change or alteration, and further that these commitments partially constitute the will. These commitments—to persons, institutions, or more abstract ideals, etc.—provide the ground upon which real *choices*, as opposed to the mere giving in to various inclinations or desires, can be advanced. "Unless a person makes choices within restrictions from which he cannot escape by merely choosing to do so, the notion of self-direction, of autonomy, cannot find a grip" (110).

This notion of finding a grip is vitally important in understanding what is at stake here. For a will to be autonomous, i.e., the will of a (moral) agent, it must itself be grounded in commitments which themselves are not subject to

easy alteration. These commitments provide the core values by reference to which all the agent's choices are made and managed, and as such represent the foundation for responsible agency itself. The agent is an agent only in virtue of these commitments, and his agency flows from them. Without them he would cease to be the person he is and would lose the very core of his personal identity and individuality.

Of course, this does not mean that an agent's core commitments are immutable and fully insulated from change. It is possible that these commitments could grow or decline. If some of them were to weaken they might be overthrown by new commitments growing out of changed circumstances. However, these changes will generally be slow and tortured; they certainly will not proceed out of capriciousness or whimsy.[2] One thing is clear, however; these changes cannot be fully comprehensive; they do not throw the agent's *entire* psychic economy into disarray.

Rather, what seems to happen is that the agent becomes conscious, in ways he previously was not, of a deep incompatibility between several of the commitments he already has and with which he is wholeheartedly identified. In other words, even in the most extreme sorts of conversions—generally religious, but not always—there always remains a kind of core, stable commitments which not only survive the conversion, but might even be

[2] I may have given the impression that the controlling factor in such changes is temporal. Clearly this is not the case; although most such deep changes take time, it would be unwise to rule out the road to Damascus possibility. Oftentimes, the most profound deep alterations in our literature concern these sorts of religious conversions. This is of course a thicket within which it would be best not to delve too far, concerning as it does, for believers, the introduction of supernatural factors (e.g., Grace and/or Divine intervention). Nonetheless, whatever the ultimate causes of these types of conversions, it should be clear that the temporal results—the volitional reorganization—are not easy or painless. Even where the conversion, like St. Paul's, is temporally quick, according to the agent's own testimony, it presented these him with a severe psychic trauma which upset his entire way of life.

vindicated by it. In the most famous cases of religious conversion, for example those of St. Paul or St. Augustine, though radically different in so many ways (blinding and sudden in the former case, slow and tortured in the latter), the agent retains a core commitment through all the changes, be it the love of God, or the love of truth (which, of course, for these two come ultimately to the same thing).

In other words, the conversion is itself predicated upon the stability of certain commitments which, in the light of changes in the agent's epistemological landscape, underwrite the radical changes in will. For instance, in Saul's case his conversion to Christianity was his response to a changed understanding of what he believed God required of him. Through his radical conversion his commitment to following the will of God remains; what changes is only his understanding of what that requires. Thus does he move from persecuting Christians to becoming one of them.

The situation with Augustine is similar in this respect. For him the changes over the course of his life, the flip-flops from a libertine life to Manichaeism and ultimately to Christianity, were, as he characterizes it, in response to the philosopher's love of truth. Confronted with problems he could not yet solve, the young Augustine moves between various commitments which seemed to him, at the time, to offer answers to his chief motivating concerns. However, even as he underwent these changes his commitment to seek out the truth remained the constant, guiding focus of his life. Whether or not he himself was correct in his ultimate analysis of what that commitment required, it is clear that subjectively this commitment remained unchanged throughout his life.

The same structural feature—the stability of a core commitment to some ideal like the truth—would hold for other, non-religious types of deep conversions as well. Copernicus' rejection of Ptolemy, and Kant's of Wolff, for instance, both answer to this description. Though some philosophers have been entirely too confident that a commitment to the truth must be the universal commitment of mankind, nonetheless I do think it is clear that some sort of commitment, whether it be to the truth or to some other object, must underlie all deep changes and conversions and remain constant through them. It is only by having this temporally extended set of commitments that anyone can be said to rise to the level of personhood and express agency. It is only against a background of stability that a change can be a "progression" and not merely a movement.[3]

Now we are in a position to see that the importance of core commitments within structure of the will itself is what actually generates the *necessity*—the phenomenon that Frankfurt calls "volitional necessity." This will help us to understand how volitional demands not arising from universalist moralities can nonetheless be volitionally normative. The necessity consists precisely in the way that such commitments determine the agent's *effective* will.

[3] When I speak of a "progression" here I do not mean that it is necessarily progress towards the good or the truth, etc. Rather, I mean it is progress in the way that a character makes "progress" through a narrative. Though some modern writers have attempted to undermine this expectation of a progressive narrative (surrealism or aburdism are examples), nonetheless the normativity of the narrative remains (and indeed provides the necessary point without which the absurdism would be meaningless. Ultimately this is connected with my understanding of a good human life to be a kind of narrative, and that person identity (partially) consists in the unity of such a personal narrative over time, but such topics are well beyond our scope here.

Frankfurt introduces his discussion of volitional necessity with an example of a mother who has decided to give up her child for adoption because she thinks, all things considered, it will be better for herself. However, when "the moment arrives for actually giving up the child… she may find that she cannot go through with it—not because she has reconsidered the matter and changed her mind but because she simply cannot bring herself to give the child away" (111). What is happening here is that she is coming up against the "*limits of* [*her*] *will*." What Frankfurt wants us to see is that it is not a question of belief or even desire—she has not changed in her belief that it would be better for her to give up the child, and perhaps she still desires (in some ways) to be free of him; rather, she cannot *will* the action. This limitation on her grows out of some core commitment within her will itself.

In this case, perhaps very much against her antenatal desires, she has come to love the child. A devotion to the child has become a central commitment for her. Thus even though she may have the conscious belief that she ought to give up the child, she nonetheless finds herself incapable of doing so, because this would do violence to her incipient love for him.

It is not that beliefs of some sort are not in play here. As the will acts it must necessarily act on and in light of beliefs about the world (and itself). Nonetheless, what this example is designed to show is that the incapacity to give up the child is not the result of an all-things-considered prudential judgment—indeed it is not a judgment *per se* at all—but rather a volitional incapacity, in which she discovers something about herself and her will (namely, that she loves the child).

Accordingly, this is experienced by her "less as a defeat than as a liberation." Although what we might call her "better judgment" is defeated by

her volitional incapacity to carry through its conclusions, nonetheless this incapacity is not an external imposition, but rather the means by which she comes to understand something important about herself. It is liberating precisely because it gives her real knowledge about what matters to her, and thus perhaps alleviates a great deal of doubt and anxiety. It is not a defeat precisely because it proceeds from her will itself—it is her "true" self breaking through, as it were.[4] The key conclusion Frankfurt is driving at is simple: for a person in a situation like this woman's, this kind of "necessity is unequivocally constitutive of his nature or essence as a volitional being" (113).

Understanding volitional necessity is vital for understanding the true nature of agency. For our purposes here we should now be in a position to see that the demands of love and loyalty are in many cases precisely a species of volitional necessity. It is in virtue of the place that concern for the objects of these demands takes within the volitional structure of the agent that he feels these demands. These loyalties rise up within the agent himself; they are not external impositions, but internal imperatives rising from his integration of his deepest loves and loyalties into his very volitional nature. These loyalties are the type on display when Robert E. Lee, offered command of the Army of the Potomac responds, "I cannot raise my hand against my birthplace, my home, my children."

[4] It is here again we recognize a key point from the previous chapter on Kant—practical reason and the will are not identical. Recall that it was Kant's contention that the pure (i.e., fully moral) will should collapse into practical reason. What Frankfurt's point about volitional necessity suggests is that this is impossible insofar as real human wills are partially constituted by their deep commitments. It is, of course, another thing to argue the moral point. Even if it is true that no real human will is, has been, or ever will be, free from these volitional necessities, this does not by itself show that it is unreasonable to require them to be. I make some greater effort to detangle these points below.

It was not merely that he favored war or Southern independence; indeed, before the war he was opposed to both. Rather, when the chips were down he discovered that he could not bring himself to act against what he perceived to be the interests of his native Virginia. I hasten to add here, that what I am interested in is not whether or not he was correct in his estimation of what his love for birthplace, home, and children required. Rather, the point at issue here is the state of his will as he contemplated acting. Though he may have favored the preservation of the Union, he simply could not bring himself to act against what he perceived to be the interests of the commonwealth he loved more. I believe this incapacity was not simply the result of weighing his preferences, but arose out of his "very essence as a volitional being."

This fatal moment is not a moment of *rational deliberation* at all, but the deepest expression of an agent's will filtered through his understanding (accurate or not) of what the best interests of what he loves requires. What is important is that the agent recognizes that the volitional necessity reflects what is centrally important to him. It would be a mistake to think that if he does not have rational grounds for his commitments that they are not his or fail to be an expression of his autonomy. With respect to his autonomy what matters is that he embrace his commitments, that he claim them for himself, as expressive of himself, whether or not he has rational grounds for holding them.

Understanding the deepest sorts of loyalty requires us to understand what is happening in cases like these. It is certainly not that all loyalties, all demands of love and loyalty, will rise to the level of volitional necessity. Nevertheless, it is important that we be aware that this sort of volitional necessity is the limit case, the embodiment of the deepest commitments in

human life. As I will argue below, it is my contention that possessing some sort of these deepest commitments is *sine qua non* for being a person and thus the ultimate limit of what any moral theory can require of us.

It is particularly important that we see that the presence of such volitional necessities constrains and informs the very possibility of rational deliberation. They represent the limits to which deliberation can come but cannot press over. Before a real, conclusive demand of love and loyalty the will stands bound, and no decision or choice can overcome its internal determination. The only way in which the will can be moved off of this volitionally necessitated course of action is for the agent to cease to love the object of concern which informs that necessity, and that is not within the bounds of rational choice. "We cannot help loving what we love, nor can we make ourselves love by a mere act of will" (114).[5]

The reason for this is the same as what we saw above. If I were able to reconstruct the nature of my will by merely choosing to do so, then my will would lack the stability necessary to make any particular configuration of the will authentically mine. Unless there is some enduring ground from which my choices proceed, as accurate reflections of what I care about, then my "choices" are little more than whimsy—they are not strictly choices at all. For a choice to be a real *choice,* it must reflect values which themselves are (to some degree) fixed and thus most authentically mine. Otherwise, they are like passing fancies and certainly do not manifest anything central to the agent's identity.

---

[5] Frankfurt continues: "The value of loving for us derives, precisely, at least in part, from the very fact that whether we love is not up to us. The importance of loving would be lost if we could love something or cease to love it merely by deciding to do so. The self-fulfillment and freedom that love provides depend upon the very necessity that love entails."

When we think hard about the nature of what we care about most—our families, our selves, the truth—I think it becomes apparent that there really is very little choice involved. I never made a choice to love my mother. This is not to say that a mature agent cannot repudiate any or all of these things; in fact, many have. It is a mistake, however, to think that because something can be repudiated by choice, that it must have been originally accepted by choice.[6]

I can choose to repudiate the color of my hair. I can choose to dye it blond, or red, or purple. But the fact that I can make that choice obviously does nothing to show that I chose for my hair to be black, or, for that matter, even assented to that fact. In the case of hair color this may be unimportant. In the case of what I discover about my will, it clearly is not.

Obviously, there are considerable differences between a physical property of my body and the constitution of my will. Nonetheless, the example does show how something can properly be said to be mine without my having chosen it. Further, the point of similarity is deeper than it might at first appear. We are so used to thinking that the will simply is the faculty of choice, that we tend to use 'will' and its cognates almost interchangeably with 'choose' and its cognates. Thus we have expressions like "fire at will" which means more or less the same as "fire whenever you choose." We speak as well of the "willing accomplice" or being forced to so something "against my will."[7] These colloquialisms notwithstanding, it should nevertheless be clear

---

[6] This is the mistake that has infected some notions of social contract theory, leading to an implausible weight being placed on consent, especially in the guise of something like Locke's notion of "tacit consent."

[7] The Stoics' idea of freedom—that my will can never be dominated except insofar as I allow it—shows the ambiguity of the notion of being forced to do something against my will. In fact, I cannot do something against my will. I can be forced to make choices I don't want to make, or act in a manner I would not have chosen otherwise, but this is rather different from actually doing something against my will.

that Frankfurt's use of 'will' as a term of art is meant to make a clear distinction between the faculty of choosing and the will.

It is important that we recognize this divergence between choosing and the determination of the will. Choosing is an exercise of practical reason. Moral philosophers have tended to concentrate almost entirely on the exercise of practical reasoning, and accordingly they have generally come to view practical reasoning as the whole story of practical normativity. Frankfurt's examples clearly show, however, that other aspects of the will are normative, aspects which may have nothing whatsoever to do with morality.

When, as we saw in the last chapter, Herman talks of the deliberative field model and the need to "normalize" one's "interests" into reasons, she is implicitly following Kant's thought that the only truly autonomous action is fully reason-guided. The question of what we should do can be (in principle, if not always in fact) settled by an appeal to the available reasons. While they would certainly acknowledge the presence of powerful desires that affect how an agent actually *does* act (even at times preventing him from acting as he would rationally choose to do without those desires), it seems as if these philosophers are not adequately sensitive to the degree to which the field of an agent's possible action is pre-determined by the higher-order commitments he has. In supposing that reasons (and thus reasoning) are basic to practical normativity, they fail to appreciate the degree to which our actions are determined by volitional structures lacking any kind of independently rational basis.

Those structures are what they are *merely* because we love what we love, and we all know that while some instances of loving may be reasonable, others are not. Ultimately, it is these loves, and the volitional necessities they

inform, which define the very scope of our practical reasoning. Contra Kant (and his myriad followers) this suggests that reasons are not always fundamental in (morally significant) agency, that deep attachments can move the will independently of the judgments they may also inform.

What I have just said here is rather radical, and liable to be misunderstood; therefore let me make clear what I am not saying. I am not saying that practical reason is unimportant or that the giving of reasons is unimportant. Further, I am not proposing that we replace the role of practical reason, within practical agency, with volitional commitment. Rather, instead of substituting one set of "rules" for another, I am suggesting that a full account of practical agency requires us to acknowledge that practical rationality is not the only modality in play in the determination of autonomous human action. Rather, it is one of potentially many volitionally constitutive elements whose interaction determines an agent's ultimate action.

What this means is that our commitment to the canons of impartial morality itself is ultimately a commitment structurally similar to our other volitionally constitutive commitments. This very surprising conclusion is endorsed by Frankfurt; he writes in reply to another essay of Herman's that his "own view is that we act morally when we are moved by love for a certain kind of world or a certain kind of life" (Buss and Overton [2002], 277). Thus our commitment to morality is founded on love—and enforced by volitional necessity—in a manner similar to our particular volitional constitutive commitments to our own specific objects of love and loyalty. The conflict that sometimes arises between the commitments then cannot be settled by

integrating them into a single deliberative field, because it is not a matter of weighing reasons—deliberating—at all.

Rather, for a given individual it is a matter of coming to grips with what most matters to him; what he loves most. Oftentimes this will be a (self-) discovery, whose precise outcome he could not have predicted beforehand. He will find out what he can and cannot do, by where the limits of his will are drawn in consequence of his deepest, constitutive volitional commitments. It may turn out that he is most committed to the ideals embodied in the principles of impartial morality and so is constrained to act in accordance with them by limits of his will. It may also turn out that other of his commitments overwhelm his commitment to the principles of impartial morality—and by the same sort of volitional necessity he is driven to accede to demands informed by these other commitments.

Of course, this working out of the limits of his will is filtered through, and informed by, the giving of reasons and the making of judgments. The agent will make judgments about what his loves require of him, and he will feel their demands play themselves out in the world. Thus the agent will generally have reasons of some sort for what he does; it is not the case that the presence of volitional necessity obviates the need for reasons or reasoning. Nonetheless, his actions are constrained by the scope of his volitionally constitutive commitments, and his deliberations occur within the boundaries set by those commitments.

Of course, this does not mean that such deep commitments are fully insulated from reflection. Agents will naturally reflect on both the compatibility of their deep commitments with each other and on what deep commitments they ought to have. This sort of reflection will not be the

everyday stuff of moral decision making, which is guided by the character one has, not one's facility at practical deliberation. Character is determined by these sorts of volitional commitments, and a good character is precisely one that is volitionally committed to the right sorts of ends (i.e., one that loves the right "kind of world" or "kind of life" whatever that may be).

One consequence of this way of thinking about practical agency is that it completely transforms many of the oldest and most problematic questions of metaethics. Perhaps the deepest question of metaethics is simply: why be moral? The sorts of answers that philosophers have sought to give to this question have tended to turn on making claims about what rationality requires, or at least what is reasonable. Indeed, as we have seen, part of the deep appeal of Kant's moral project is his promise to show how morality is the inescapable consequence of rationality and thus to be avoided only on pain of irrationality.

However, these sorts of answers to the question—whether Kant's or other attempts to show the reasonableness of morality—have ultimately failed to convince. Their claims to ground morality in what is reasonable have floundered on the shoals of skepticism, both epistemological and moral. This skeptical question undermines the moral project precisely because it generates a false expectation of what sort of answer will be adequate. This challenge will not be met with an argument whose conclusion flows apodictically from the pure tenets of reason alone.

What Frankfurt's analysis of practical agency suggests is that the challenge will be met only by pointing out the strength of our commitment to morality—namely, that we love a certain moral ideal. Moral skepticism, as he says, "is not a problem of truth. It is a problem about confidence" (Buss and

Overton, 277). Thus the only "justification" of morality we can give to the moral skeptic is to show him that and *how much* we love it. The *why* is lost, and the demand for a conclusive reason to be moral is as inappropriate as the demand for a justification of my love for my mother or my wife. At some point the loves which partially make up who I am just *are*, and have no further justification than that I am who I am.[8]

I suspect many will find this deeply unsatisfying. The natural rejoinder is to this is to observe that even if what we love does somehow determine the ultimate limits of the will, nonetheless one might observe that the question of what we *should* love remains unanswered. Thus, even if love is a (perhaps, *the*) centrally important determinant of the structure of the will and thus what we actually do, nonetheless the same moral questions can now be reformulated into questions about what volitionally constitutive commitments we ought to have.

It is a mistake, however, to think that the question about what we ought to love is the same sort question that philosophers have traditionally understood when trying to understand what we ought to do. The reason for

---

[8] Frankfurt is especially revealing in the following passage:

> Many philosophers and other people think it is important to find ways of demonstrating conclusively such things as that it is reasonable to tell the truth, and to refrain from murdering people, and from inflicting pain gratuitously. They are not satisfied with the fact that many of us are moved lovingly—as a matter not of choice but of volitional necessity—by ideals of straightforwardness and of consideration for others.
>
> But why are they not satisfied by this? If someone tries to interfere with my children, I will try to fight him off. If someone tries to interfere with our efforts to bring about or to sustain the kind of world we love, why should we not be as confidently disposed to fight him off? We do not need any knockdown reason for knocking him down, other than the reason that he is trying to damage something we love (Buss and Overton [2002], 277).

this, of course, is that what we love is not fully up to us, while, at least as traditionally understood, what we do is. To a large degree we are passive with respect to what we love; we discover those loves as facts about ourselves and thus come to observe the shapes of our own wills. The deepest and best loves are essentially a *response* to the value of the most lovable (i.e., the best) of objects.[9]

Of course, it should also be pointed out that the limits established by volitional necessity are just that—limits. These limits will not always, or even commonly, be challenged. Furthermore, for the vast majority of people a love of the truth and "ideals of straightforwardness and of consideration for others" will also be volitionally constitutive, and living up to those ideals will generate more or less all of the practice of morality we would want.

In other words, volitional necessities set the limits within which the conventional deliberations of morality take place (for a given agent), and do not always or even commonly override them. In fact, the volitional necessity arising from the agent's commitment to those principles will itself determine his compliance with them. As such, the demands of love and loyalty will not simply run rough shod over all the conventional principles of morality, precisely because the wills which are partially constituted by commitments to their own love and loyalties are also partially constituted by a commitment to those very principles. The main point, echoing Williams, is that the commitment to the principle of morality will not necessarily be always

[9] This point about the receptivity of love is a very old view, of course, and one apparently advanced in several Platonic dialogues, e.g., the *Phaedrus*. My central concern here is the draw the deep distinction between the way love operates in the will and the way so many philosophers have conceived of practical reason. It is certainly not the case that loves are permanently fixed, they can grow, fade, and be cultivated. However, they are certainly more resistant to change, the express that resistance through volitional necessities, than the reasons which inform the deliberations of practical reason.

overriding and will sometimes be rejected by an agent whose will is shaped by love.

It may help if we consider this under a slightly different description. Generally it is held that *ought* implies *can*. In other words, it cannot be a moral requirement to do the impossible. For there to a binding *ought*, it must be possible to achieve its ends; otherwise, the *ought* is vacuous. Of course, this is exactly what happens with questions of what we ought to love. Sometimes it may be impossible—volitionally impossible—for me to love some object, or perhaps more pressingly, not to love some person or some thing that I do. I simply am not completely free with respect to my loving.

Nonetheless, there is more wiggle room here than I might have seemed to imply. Frankfurt himself notes that the fact that one cannot "rigorously demonstrate which ideals are" proper to have "should not be taken to entail… that our volitional necessities must merely be acknowledged as givens—that is, accepted passively as brute facts with respect to which deliberation and rational critique have no place" (Frankfurt [1999], 116). While it is true that I cannot simply choose to love or not to love, I can make choices which will affect my loving. In other words, I can potentially set out to change the conditions in which my will is forced to operate, and in so doing what I care about may change over time. These choices will often be the result of a rational examination of what is entailed by our having various moral ideals—of "a certain kind of world or a certain kind of life"—to which we are centrally committed.[10]

[10] This is analogous to somebody afflicted with an unrequited love. Such a person might decide that he needs to avoid places and situations that remind him of his beloved (and act on that decision). Over time we would expect that his love would change accordingly. That is not the same as deciding to stop loving, however.

Some of our examples above answer exactly to this description. It was by coming to realize how his existing volitional commitment to the truth led to God that Augustine came to love God. That commitment, in turn, grew to encompass more and more of his will until it overcame everything else, so that he saw all of his commitments through the lens of his commitment to God. Again, whether he was right or not about the nature of these commitments is not important; the fact that they did arise and completely determined the nature and scope of his agency is.

Thus it can, and will, come to pass that we are required to examine how our ideals interrelate and interact and how they may be potentially in conflict. When they are, I may be forced to give up one or the other of them in response to this conflict and which of them I give up will manifest my true character. When the requirements of an impartial moral ideal to which I am committed conflict deeply with the requirements of my love for some other object what I ultimately do will reveal what I love more.

As Williams implied in the quote with which we began, it is not simply a matter of justification as conventionally understood at all. Rather it is the case that for a will to be committed to anything at all, it must, to some degree, be committed to objects that are particular to it—objects of love and loyalty.[11] Once we see that it is the *love* of the world or way of life informed by principles of morality which underwrites the agent's moral activity we realize that those principles cannot be fully regulative of that love. They can shape and guide action, in just the same way that the conflicting demands of any

[11] The particularity of these objects does not preclude other agents from possessing similar commitments, just as my siblings also love my mother. However, without going into anything silly like a comparison of our respective loves for her, I think we can still say that the place in each of our wills that our love for her takes has unique content of various sorts based on our differing experiences, etc.

objects of the agent's love can require him to rethink and modify his actions and beliefs with respect to them. However, such principles are generally not the agent's ultimate commitment, because his agency itself is founded on special commitments that make him who he is, and make his life valuable for what it is. He can of course share his commitment to morality with others, but his agency arises from the particularity of the constitution of his own will.[12]

I think we are now in a position to appreciate why the demands of love and loyalty are directed at values which cannot be accounted for within the framework of the universalist theories. These values are precisely those which find expression in the types of love which eventuate in volitional necessities. To possess these values is precisely to love in a certain of way—wholeheartedly, disinterestedly, and without reservation.[13] It is a love directed above all at the good of the object of that love.

It is presumably true that one could love the demands of impartial morality in a similar way, such that one could be driven by volitional necessities to act as a universalist theory might demand. It is not my purpose to deny that possibility. Rather my point is that while this analysis of volitional necessity can encompass those demands of morality within a certain kind of demand of love, the converse is not true. The proponents of universalist moral theories cannot accord the proper value to the demands of love and loyalty.

---

[12] It is possible to have an agent most committed (i.e., in love with) a world informed by the principles of some abstract, impartial morality alone, but such individuals would probably seem inhuman, devoid of mercy and sympathy—Lycurgus instead of Solomon.

[13] "To say that a lover is disinterested means simply that he desire the good of his beloved for its own sake rather than for the sake of anything else" (Frankfurt [1999], 168).

This deepest sort of love is as valuable as it is precisely because it is not subject to the same sorts of justificatory requirements as other values. Its value inheres in the fact that it is so fully expressive of the agent's will that he himself cannot escape from it. It is precisely this relative immutability—this stability in the face of skepticism—which makes these values the core of the agent's being. Without them, his agency dissolves away. It is precisely his confidence in his love for what he loves that gives him the foundation on which to stand, which makes for a life worth living.

This is why the ultimate issue is not the knowability of reasons for acting, but the confidence that an agent has in his own volitional commitments. Deep skepticism undermines agency precisely because it elevates an epistemic difficulty into the will. It vitiates the agent's capacity to act by making him believe that all of his actions have to have particular kind of justification in fully grounded reasons. The skeptic's inability to act thus reveals a kind of inability to love in the right sort of way, and so reveals a defective will (at least with respect to certain values, though not necessarily all).

The deepest demands of love and loyalty are the expressions of a self-constituting love. They press beyond *conventional* questions of moral justification. Again, however, this does not mean that they have to be beyond examination, but it does mean that a deep skepticism about them can undermine them to such a degree that they cannot survive in the same form. In other words, to demand justification—to believe that it is appropriate to seek a moral accounting of the deepest convictions of love—is already to devalue them. It is to fail to see how their value inheres in precisely the depth

of the agent's commitment to their objects. This truly is "one thought too many."

(Re-) Connections

Before closing this chapter, I thought it might be useful to connect some of the foregoing observations to some of the concerns raised in the limited literature on loyalty that we examined in Chapter II. Specifically, I want to draw connections with my discussion of the two central figures in the philosophical examinations of loyalty we considered in Chapter II: Andrew Oldenquist's and Josiah Royce's.

Recall that in his article "Loyalties," Oldenquist made the following suggestive claim: in "terms of the logic of the reasons they provide, loyalties are a third category of the normative, distinct from both self-interest and impersonal morality…. Loyalties are part of what make our societal worlds go around, but they cannot be understood if we try to turn them into either impersonal duties or sophisticated egoism" (Oldenquist [1982], 176). Further, "our wide and narrow loyalties define moral communities or domains within which we are willing to universalize moral judgments, treat equals equally, protect the common good, and in other ways adopt the familiar machinery of impersonal morality" (*Ibid*., 177).

Fascinating as these suggestions are, Oldenquist did not adequately develop and substantiate these claims. However, their similarity to claims I have made arising out of the Frankfurtian analysis is hardly coincidental. As Oldenquist understood it, loyalties represented a separate "category of the normative." When rearticulated in terms of volitional necessities we can see

how this is possible. Oldenquist clearly recognized how deep loyalties could become normative without being in any way derivative from duties arising from impersonal morality. We see that this is possible insofar as loyalties are a response to commitments which partially constitute the will.[14]

The *normativity* of these volitionally constitutive commitments is very important. There is a sort of general presumption that the only significant normativity is *moral* normativity. The norms of universal morality are opposed to the contingency of desire or whim, etc. as fundamentally more serious and more important. What we need to acknowledge is that the normativity of morality within the will has the same status as the normativity of any other deeply held, volitionally constitutive commitments. Thus, deep attachments and loyalties are just as normative as the requirements of impersonal morality (which is, insofar as it is effective within the will, just another a form of love as well).

Oldenquist's second claim—that loyalties define the scope of impersonal morality—is strikingly reminiscent of Frankfurt's account of how volitional necessities define the limits of the will. Again Oldenquist's arguments are reaching in the proper direction and adumbrating the right conclusions, but it is through the application of the Frankfurtian model that we can see how the conclusion can be supported. Oldenquist clearly was onto something important, and the congruence of his claims with the consequences of taking seriously the importance of love in the will provides further support for it.

---

[14] Of course Oldenquist speaks of loyalties in terms of the "reasons they provide," while I have claimed that to some degree reasons are transcended by volitional necessities.

There are similar resonances with Royce. Recall that in the *Philosophy of Loyalty* Royce claimed that loyalty, "fixes our attention upon some cause, bids us look without ourselves to see what this unified cause is, shows us thus some one plan of action, and then says to us, 'In this cause is your life, your will, your opportunity, your fulfillment'" (Royce [1908], 21). What he seems to imply here is that it is precisely through a loyal commitment to some "superpersonal" cause that I come to have a self. Thus, the object of my loyal commitment is not external to me, but actually constitutive of my will. In other words, it is only insofar as I am able to commit myself loyally to some cause that I even have a will at all. Therefore, when I act in accordance with its demands it is not as if I am acted upon by something external to my will, but rather I fulfill my own willing loyal commitment to a cause. "[U]nless you can find some sort of loyalty, you cannot find unity and peace in your active living" (23). Thus loyalty (to a cause) is what binds one's life into a coherent whole—that makes it a life.

Royce attempts to sustain this claim with a neo-Hegelian metaphysics which need not detain us here. Rather my interest is the striking congruence of Royce's observation with the claims I have advanced out of the Frankfurtian model, specifically the connection with the claim that core commitments are partially constitutive of the will. The key point of connection is the idea that what it is to have a self is to have deep commitments.

Royce even goes so far as to claim that the particular object of one's loyalty is relatively unimportant at least insofar as it does not conflict with other loyalties, i.e., is not "destructive of loyalty in the world of my fellows" (Royce, 56). The value of loyalty for Royce is not dependent on the value of

that to which one is loyal. Rather loyalty itself is valuable in itself, precisely because it is the means by which I come to be the self that I am.

This again closely tracks some conclusions Frankfurt draws about love. "Loving is valuable in itself, and not only in virtue of the value of its objects. Other things being equal, our lives would be worse without it" (Frankfurt [1999], 173). This is because the very "act" of loving represents a growth in the lover. His life becomes more valuable in consequence of his loving; "without loving in one or more of its several modes life for us would be intolerably unshaped and empty" (*Ibid*., 174).

I certainly do not expect to have proven my case with these appeals to Royce and Oldenquist. Rather, what I hope to have done is much more modest. My intention was to draw connections between the analysis of the demands of love and loyalty and independent claims advanced by very different thinkers coming at these problems with very different backgrounds, inclinations, metaphysical commitments, and philosophical methodologies. Nonetheless, they arrive at importantly similar conclusions. Obviously, my hope is that this unity is derivative from the unity of the truth about the phenomena under examination. I suspect that these similarities are anything but coincidental.

Conclusion

In this chapter I have tried to lay out an analysis, drawn and adapted from Harry Frankfurt, which provides a framework for understanding the demands of love and loyalty within a sophisticated understanding of moral psychology. I hope to have shown how the demands of loyalty grow out of

volitional structures which are themselves the result of possessing a certain kind of commitment or love. I hope also to have shown how the values that inform these demands of love and loyalty are deeply embedded within the nature of the will itself and must be understood in those terms.

In the next chapter, I turn my attention in a different direction, toward the question of how a will—a self—comes to be. My concern with development grows out of the recognition that if agency flows from our deepest volitional commitments, and if those commitments cannot all simply be chosen, then we need to understand where they came from in order to have a handle on the nature of human agency. Later, in Chapter VII, I will utilize the model I have adapted from Frankfurt in this chapter, along with the developmental story I will describe next, to show why particular loyalties are important to being a certain sort of self and why those loyalties are an inescapable pre-condition for having a life with "substance" expressive of a human-shaped will of the sort we value.

CHAPTER VI

A Developmental Account of Moral Psychology

In the last chapter I tried to show how the demands of love and loyalty could manifest volitional necessities directly reflecting structural features of the will. In this chapter I will move in a different direction. Much of my discussion thus far has focused on what I have characterized as *unchosen* loyalties—the demands of love and loyalty arising from commitments to objects such as one's parents, children, community, country, etc. I now want to ask: if the commitments that underlie these demands are unchosen, whence did they come? It is in attempting to give an answer to this question that I believe we will make some progress towards understanding the sense in which these sorts commitments are necessary components of a human life with "substance."

Lear

To begin answering the question whence unchosen loyalties come I am going to introduce a discussion by Jonathan Lear which offers us some helpful tools. I wish to state quite clearly from the beginning that my project is very different from Lear's and there are many facets of his argument from which I

demur; nonetheless, I think his concern with Freudian developmental *empirical* psychology points the way towards giving an account of developmental *moral* psychology. My hope is that in articulating the features of the development of moral psychology we may come to better recognize the central role loyalties, especially to unchosen objects, play within our wills.

Lear is a psychoanalyst as well as a philosopher, and he has sought to inform the philosophers' debates with insights he claims from Freudian psychoanalysis. Lear claims that there is a deep continuity between the Socratic philosophical project and Freudian psychoanalysis.[192] In a fascinating little book, *Love and Its Place in Nature*, Lear wants to introduce a Freudian conception of love into philosophy, especially as it relates to the way in which egos develop, and in so doing he expects that this will transform several important questions in moral psychology.

> Modern philosophy in general has a tendency to investigate the structure of the human soul in isolation from how that structure came about. The thought is that philosophy is only concerned with what the structure of the soul is: how this structure developed is a topic for psychology. The problem with this reasoning is that if the individual is by nature a response to a good enough world, we may be unable to determine the structure of his soul is if we remain ignorant of the conditions in which it came into being. (Lear [1990], 186)

It is only by asking how the soul, or will,[193] develops that we can come to understand what it essentially is. If we recognize that the mature will is in

---

[192] Lear writes: "Psychoanalysis" is a word which, even without our translating it, Socrates would have understood. In fact, Socrates invented philosophy by devising his own peculiar brand of psychoanalysis. Through an examination of one's own soul, Socrates believed, the truth would emerge and it would set one free. (Lear [1999], 183)

[193] Lear seems to freely interchange 'soul,' 'ego,' and 'I'; I will continue this, sometimes throwing in 'will' for good measure. This is, of course, not to claim that all of these terms have precisely coextensive referents. I hope context will supply any needed disambiguation.

part a reaction to the world in which it comes into being and finds its shape, we can begin to see how important it is to understand the relationship of the internal structure of the will to the external structure of the world. If the structure of the will is somehow in part a reflection of the structure of the world, then we can begin to understand how loyalty to those structures can be an authentic expression of that will. Before we can get there, however, I want to examine the Freudian groundwork of Lear's discussion from which I will then borrow important concepts to motivate my own significantly different project.

Before I begin let me reiterate that it is not my intention to take up Lear's account of Freudian psychology wholesale. I can remain somewhat agnostic towards its accuracy while nonetheless borrowing some of the fruitful concepts Lear develops out of it. As I indicated above, my goal is not a developmental *empirical* psychology, but a developmental *moral* psychology. Chronologically, within the life of a given individual, the development of a *moral* psychology is presumably subsequent to the first stirrings of the psyche. Of course, the latter is the necessary ground for the former and will naturally affect it. Nonetheless, in taking up some of Lear's arguments I believe my purposes in investigating the unambiguously moral aspects of development can be served independently of his more strictly psychological (and, perhaps, pre-moral) purposes.

Lear introduces an important distinction I find very useful.

> The attitude of *holding someone responsible* is one that one can take toward oneself or others. In holding oneself responsible one is essentially taking a third-personal stance with respect to oneself and deciding whether one's character, actions, thoughts or feelings are worthy of praise or blame…. Accepting

> responsibility is essentially a first-person relation. In accepting responsibility I acknowledge who or what I am. (Lear, 66-7).

This distinction essentially involves the attitude one takes towards certain aspects of oneself. As Lear explicitly characterizes it here, holding oneself responsible is an attitude of appraisal—deciding whether something about oneself is "worthy of praise or blame." In accepting responsibility, on the other hand, what is at stake is not a judgment of the worth or value of some aspect of oneself, but rather one's ownership of it. In other words, to hold responsible is inevitably to take a kind of external stance, even if one is doing this towards oneself, whereby one judges something or someone against some external standard. To take responsibility is to firmly stake a claim of ownership—to assert the *mineness* of some action or property.

Lear introduces this distinction in the context of his discussion of Freud's treatment of a patient who represses certain emotions by holding herself responsible for them and instead of taking responsibility for them as her own. However, this distinction comes up again in a more abstract discussion towards the end of the book of what he calls the "other Oedipus complex." The question is why Oedipus, whose fate seems to have been determined for him before he was born, ultimately accepts responsibility for his actions. Lear writes that Oedipus:

> …in accepting responsibility for his acts, is claiming that the truth that ultimately matters is, as he says of his blinding, that "I have done it by my own hand." Whatever the gods ordained, Oedipus says, the fact is that I did it. Oedipus is in effect claiming a part of nature for himself. He treats himself as a locus of activity and in that way demands to be distinguished from the rest of nature…. Now, Oedipus may also *hold himself responsible*: for he may take up an observational stance with respect to himself and judge his acts to be praiseworthy or blameworthy. In holding himself responsible, Oedipus is claiming to be a fit

> subject of ethical or aesthetic evaluation…. In accepting responsibility, Oedipus is making an altogether more elemental claim: "these acts are *my acts*." Oedipus thus constitutes himself as an agent, a locus of activity. (Lear [1990], 170-1).

There is a lot going on here, but what I want to focus on now is the notion that it is by taking responsibility for something which he did not choose that Oedipus makes himself "a locus of activity." It is this act of claiming responsibility for something which he himself did not make—of something that pre-exists—which gives substance to his agency. For it is in this claim—of asserting that these action are *his*—that he manifests himself a cause in the world (a "locus of activity") and not a mere plaything of the gods.

Here Lear is filling out his fundamental claim that a soul develops in response to the world. The internal structure of the soul is a reflection, through the psychological act of identification, of the world, which it "devours."

> Psychic structure, Freud realizes, is created by a dialectic of love and loss. The structure of the mind is an inner recreation of the structure of the loved world. Mental structure develops with the infant's increasing appreciation that the loved world exists independently of him and is not immediately responsive to his wishes. (160).

The developmental account here is meant to explain something extremely mysterious—namely, how a pre-psychological entity (an infant) comes to possess a genuine human psyche. The infant begins in an undifferentiated state, unable to recognize his own independence from the world. However, the world turns out to be not perfectly responsive to the infant's needs and wishes and out "of the ensuing frustration and disappointment, I am born."

With a suitably attentive mother an infant's needs will be almost instantly satisfied. And yet, no matter how sensitive his mother is, at some point she will fail to feed him exactly when he hungers, and he will begin to perceive that his hunger does not automatically call forth the "good breast" (as Melanie Klein called it) to satisfy that hunger. It is in these very first frustrations that he will begin to perceive his independence from the world and thus begin to differentiate himself from it.

The process continues as in response his "recognition" of his independence from the external world the infant begins to invest the world with a very basic form of love—identification. By identifying with parts of the world the proto-psyche attempts to pull the receding world into itself. The world is "receding" insofar as its non-compliance with his desires manifests its independence from them. The structure of that world becomes written into the structure of the developing psyche through this psychological process of identification. It is key to see that this identification is a *psychological* process—"not merely a neurophysiological" one; it is active, not passive. It is out of the "swirl of drive activity" (the most basic psychological drives which Freud called "primary process") that this development occurs. However, it is inevitable that this primary process, "from the perspective of the already developed I, looks like magic" (Lear, 163).[194] That is, from our present

[194] In all of these discussion about the psychology of infants, etc. there is complication of which we must be aware. There is surely a sense in which the infant, at his stage, cannot possible *recognize* anything. Not only does the infant have no precedent for this new fact, he has no precedent for the process of recognition itself. And yet, somehow or other this, or something like it, must come to pass. So it may be that we must always speak loosely or metaphorically as we attribute complex psychological states and processes to nascent psyche which cannot have possibly expressed them. The whole point of this sort of bootstrapping argument is that it is in the struggle of the nascent psyche to achieve this complexity that the complexity comes to be. In the end, though, it may necessarily be *magical* from our perspective, where *magical* here means that we are not only unaware of the processes that underwrite it, but may

perspective, looking back on our own development, it appears almost mystical as something recognizably like a mind seems to come out of something that looks, at best, merely animal.

Of course not just any world will facilitate this process of identification. It must be a "good-enough world," where the infant's physical needs are sufficiently met, but just as importantly his psychological needs are satisfied as well. In a world that is unresponsive to the infant's developing psyche, he may be permanently stunted. The responsiveness necessary is not only, or even most importantly, the care of his physical needs. His need for love is just as pronounced, and perhaps even more pressing, than his need for milk.[195]

What is required is that the world of the infant appear continually more complicated and sophisticated to him; it is the proto-psyche's attempt to "metabolize" this increasingly complicated world, though identification, that eventuates in its own increasing internal complexity. Furthermore, perhaps the most important part of that complicated external world is the love it showers on the infant by way of his mother. It is maternal love which drives the infant's developing psyche as that psyche is caught in a process where it attempts to take in that love and respond to it.

This "dialectic of love and loss" (love and frustration) motivates the increasing sophistication of the nascent psyche—the interplay between the less developed soul of the child and the mature soul of his mother, as she lavishes her love and care upon him. The child's soul literally comes to be by

---

necessarily unable to grasp the essential nature of that development *from the inside* (this despite the fact that everyone of us must have gone through it).

[195] The classic examples are those children warehoused in orphanages in the old communist block countries (especially Romania), whose physical needs were minimally supplied, but whose need for love was tragically unmet. These children, when revealed later, were psychologically much "younger" than their chronological ages, and radically undeveloped intellectually and emotionally.

taking into itself the love it receives from the world, by actively identifying with that world and thus reconstituting within itself the structure of that world as lovable and loving. Love here is a fundamental psychological force, the most fundamental psychological force and indeed underwrites the very existence of the soul.

> Love… fuels a dialectic of development. Psychic structure can continue to develop because the world outstrips my ability to appreciate it. As I develop in complexity, so does the world as it exists for me. The internalization of structure can thus continue at ever higher levels of complexity and refinement. (177).

At this point, having summarized Lear's very suggestive account of psychological development, I want to take a step back. As I indicated at the outset of this discussion, my true interest is not in the earliest steps of psychological development, but in the development of a moral psyche. Obviously, this happens after the kinds of processes Lear is describing, but as his own appeal to the Oedipus case suggests there are fundamentally similar issues and concepts in play in the development of the more mature stages of a psyche that are proper to moral psychology and moral philosophy.

One part of Lear's discussion I have under-emphasized (compared to the role it plays in his exposition) is the importance for Freudians of primary process—the "swirl of drive activity" which is the earliest manifestation of the (proto-) psyche. On Freud's understanding the drives that make up primary process continue to exist and make themselves felt even after the development of a mature psyche. Indeed, a great part of Freudian psychoanalysis consists precisely in the integration of these drives (the id) into the analysand's

conscious life.[196] It is in this context that the Oedipus case comes into play, in that the "second Oedipus complex" consists in taking responsibility for this drive activity and integrating it into one's life.

It is precisely here that I break with Lear's project and turn my attention in a different direction. My interest is not in the way in which such primal drives are integrated into the agent's life, but in what I see as the analogous way in which the sources of value the emerging agent finds in the world are written into his developing moral consciousness. In other words, the powerful suggestion of Lear's that I want to follow is this signal insight that what the soul is, it is precisely because of the manner in which it has "metabolized" the sources of value in the world around it, constructing its own structure in light of and in response to these values. I think this process is again driven by love, but a love more sophisticated than the psychological process of simple identification (namely, a kind of loyalty).

In the last chapter I noted Frankfurt's thought that autonomy itself must find a grip in the fact of the agent's limitation within "restrictions from which he cannot escape by merely choosing to do so." One obvious question, which Frankfurt never addresses, is from where these restrictions come. I think a lot of the resistance to this notion comes from a suspicion that the sorts of restrictions in play—founded in unchosen volitional necessities—are external or illegitimate impositions. To a certain mindset used to the celebration of freedom and choice, the idea that there are these sorts of

[196] Of course, the integration of the super-ego is also part of the goal. Freudians have often understood mental health to consist precisely in the successful integration of the id and super-ego into ego, the unification of all the parts of the soul into the agent's conscious understanding of himself.

inescapable restrictions, and further that autonomy itself is grounded in them, is difficult to accept.

My hope is that understanding how a moral psyche develops will alleviate some of these concerns by showing how such restrictions are the unavoidable consequence of the advent and growth of moral consciousness. As I have already indicated, I believe that this process works in a manner analogous to that discussed by Lear. For moral consciousness to develop it must somehow internalize values from the external world, and I believe it does this precisely by a process of taking responsibility for the unchosen.

One of the most important things to notice about this developmental process is that for the process to be successful the world must be (at least partially) accommodating. The world must make a place for the nascent personality to develop into. Like the earlier dialectic described by Lear, this dialectic between the developing agent and the world requires the world to appear continually more complicated and challenging. At first, of course, the *world* in question is, at least for the child, little more than his parents (even just his mother). Gradually, however, the world, including both its expectations and rewards, begins to encompass a larger field.

When the child is born the accommodation afforded to him is primarily formal. Society takes note of the child and makes a space for him by immediately recognizing his personhood in law and custom. In gaining a name, and a birth certificate, the child is given legal personality and the first and central subset of his rights in positive law. Of course in reality this a kind of legal fiction, but it is an extremely important fiction.[197]

[197] Of course my description of this process partakes of the details of our contemporary western world. However, I do not think the details (birth certificates, etc.) are particularly important here. Rather what is important is that accommodation is made in some way.

As some consequentialists have argued it seems almost perverse to grant rights to a newborn. It is difficult to imagine what external functionalist criteria of personhood (e.g., rational activity, etc.) a newborn could possibly meet. Nonetheless, society rightly makes this gift to the newborn of legal personality, not because he is capable of exercising it, but in part precisely because he is not. This gift is the first step in a long process by which the external world makes a place for the child to grow into, by giving to him more than he can handle and making demands on him that he cannot as yet meet.

As the child ages and gains new competencies over himself and the world around him, the accommodation takes the form of particular roles which are opened to him. This is the everyday stuff of parenting; as the child ages the parent is forced to accommodate his increasing personality. He is gradually given more lee-way and more choices. Indeed, the child will demand more freedom and responsibility over himself and the world around him, and the wise parent will of course use the child's desire for independence to help him grow. But this is and must be a gradual process. Nonetheless, as new roles and opportunities are opened to the child the world also begins to make new demands upon him.

Some of the very first of these demands are of course bodily self-control. The child is instructed that his body is under his will, and indeed one of the first steps in the development of the will is precisely the bending of his body's voluntary processes to it. What is key, however, is how new demands

---

Whatever the particularities of a given society's customs and traditions, etc. may be, I think it must be a universal practice of mankind that some sort of accommodation is made, even if at first it is only the parents or even the mother who makes it. Additionally, I will speak of mothers and parents, but of course what is vital is not that the adult "care-givers" be biologically related to the child, but that they invest him with their love. Generally parents will do this better than others, but this obviously need not be the case.

are gradually placed on the child by the world. As he ages the world begins to *hold him responsible* for meeting its demands. The world, in both his family and in society at large, demands of the child that he step up to new responsibilities not of his own making and makes him subject to both praise and punishment in accordance with his responses.

The beginning of this process is *training*, indeed it is close to conditioning. Demands are made, and enforced by the world (generally by parents). In our society the enforcement can take the form of punishment, but mostly it is a sort of encouragement, which is to say, an offer of (more demonstrative) love. This can seem (and be) a rather mechanical process, similar to training an animal (e.g. housebreaking a puppy). Over time, however, and in all societies that produce morally mature agents, the nature of the process must fundamentally change from training to *education*.

This distinction between training and education is vitally important. The demands the world makes in the process of education require of the child not merely that he meet those demands mechanically, but that he begin to understand that he is being held responsible for meeting them. To the extent that moral education is successful, the child acts not out of a desire to avoid punishment but from values taken up and internalized. They are written into the child's will by his coming to care about those values. At first, this caring is filtered through his love for his parents and his attempt to emulate their commitment to these values, but over time I believe a significant transition occurs. The child begins to see himself not merely in the context of what is demanded of him; rather the values the world imposes on him eventually become available to him first-personally. This transition occurs precisely to the

extent that he begins to *accept responsibility* for the roles in which he finds himself and the demands the world makes on him.

In other words, the development of a moral psychology proceeds along the same sorts of lines as Lear's description of the "second Oedipus complex." The values which society imposes by holding the developing child responsible become embedded in his will by his coming to care about them: at first concomitant with his loving of his parents, etc., but eventually for their own sakes. These demands are external to him and are imposed on him. Over time, however, the values to which these demands are responses are directly written into his will as he comes to care about the ideals which they represent. As such, these values become *his*, and he is rightly able to claim ownership of them, even though he received them from the world. What matters in making them his is not that they be entirely unique to him, of his own original composition, but that they be *willed* by him, as expressive of some of the central commitments that constitute his will.

My narrative has moved quite quickly and may leave a mistaken impression that I think this process is easier or faster than any parent knows it to be. Of course the changes I am describing take place slowly over the course of years. Additionally, different individuals respond differently and develop at different rates and to different degrees. However, underlying the particularities of any given child's development is the universal necessity that this process take place to some degree or another. For it is this transition from being held responsible by others to accepting responsibility for the demands and duties entailed by various ideals, moral and otherwise, that is the very hallmark of maturation. It is this maturation that is *sine qua non* for the achievement of agency and genuine autonomy.

My central claim here is homologous to Lear's but applied to a different (and more morally significant) stage of life: the mature will (i.e., the ground of moral agency) develops out of a developmental response to a good-enough world. The will comes to be precisely by claiming for itself, through accepting responsibility for it, a part of the world not of its own making. New social roles, with their various demands and privileges, gradually open to the child, and as he takes them up and comes to identify with them he pulls into his own will a reflection of the values they entail.

This need not be, and most often is not, a conscious process. Neither, of course, is it mechanical. It proceeds not through simple identification but rather through the more sophisticated modalities of loving. These include the embracing of high and romantic ideals and in seeing that even when society fails to live up to those ideals they remain powerful within the wills of those who commit themselves to them. Further, these various lovings are mutually intertwined, and love for one sort of thing (e.g., a person) often leads naturally to loving a different sort of thing (e.g., beloved's ideals). Thus do successful parents pass on their values and traditions, not by imprinting them, but by cultivating a love for them in their children.

The good-enough world is precisely the world that both offers appropriate opportunities and makes appropriate demands. The demands (the world's holding responsible) facilitate the opportunities (the child's taking responsibility). A world that is insufficiently demanding will naturally delay or even prevent this process, which is to say, delay or prevent the development of genuine selves. Thus the world cannot be too responsive, else the child will lack the kind of frustrations which develop character and resolve. At the same time the world cannot be too recalcitrant or the child may

lack the opportunities he needs to find his own individuality and personal responsibility.

Recall that the Frankfurtian model of the will holds that the will is itself significantly constituted by its deepest commitments. Who we are is literally determined by what and how we love. If this is the correct understanding of what it is to have a self, a fuller understanding of who we are requires an understanding of how we come to love what it is that we come to love.

My suggestion here has been that human beings grow out of a loving response to a loving world. We love because we must. To some important degree we love what we love because the world in which we are raised writes its values into us by articulating the demands to which we are held accountable and for which we each must ultimately take responsibility. As such, the sort of volitionally constitutive central commitments which ground our most important projects, including and especially our commitments to morality, are precisely the ones taken up and internalized from the external world in this way.

This is different from a kind of training model that depends only on positive and negative reinforcement (reward and punishment). While rewards and punishments may be useful at first, especially for a young child, in the end they will not shape the soul through love in the way I think necessary to underwrite a morally mature agency. While such methods might eventuate in someone who refrains from doing the "wrong" thing, such an agent will not be motivated by his love for the values (moral and otherwise) which give meaning and structure to his world. As such he will be more of an automaton than a genuine agent.

Of course many people never fully manage moral maturity. The psychopath is quite simply an individual who has never taken responsibility for what is demanded of him by the values his society embodies.[198] Even some who do not become criminals nonetheless fail to make the full transition. Their pathological tendencies are kept in check by external force or the successful threat of external force. But the truly socialized human being, the mature moral agent, has succeeded in internalizing his society's values, precisely because he has taken them up and made them his own by taking responsibility for them. It is not just that he externally conforms to what they require, but that he expresses them through his very will, in how he lives his life and conceives of his commitments to those around him.

This means that a self cannot be made; personality cannot be mechanically imposed. Ultimately, the mature agent cannot not merely comport himself externally to society's demands, he must take them up as his own. He must claim value out of the external world by embracing those values—as embodied in his care for specific persons, institutions, ideals, etc.—as the objects of his loving. It is only by doing so that he finds a self. Paradoxically, it is also by doing so that he himself becomes a source of new value as an authentic bearer of social value within his will, constituted by his higher-order commitment to them.

It should now be clear why the self cannot be self-created. The self must develop in reaction to and in a relation with an accommodating world. The will develops by loving the world and thus partially recreating the world's structure within itself. It should also be clear that this process is, to a large

[198] This is true whether the cause of this failure is something for which he can be rightly held responsible or not. Mental illness may prevent his successful accepting of responsibility, but so may an overly permissive or otherwise defective community and/or polity.

degree, unchosen. Because what is at issue is the development of the will itself, it cannot be the case that the will is choosing what it loves. Rather what it finds itself loving begins to establish the contours of what choices are available to it as it matures.

Frankfurt's conceptual point, that autonomy requires necessity, is another way of saying that the will cannot choose its own structure, because for a choice to be a genuine choice it must be a reflection of the pre-existing structure of that will. The developmental account I am offering here attempts to explain how the ground of autonomy—a stable pre-existing structure of the will—must come out of the unchosen loving wherein the developing will takes in the values of the world. The structure of the will must develop as a reaction to and in a relation with the structure of the external world. The concrete manifestation of that external structure are the people and institutions within it and the ways in which they express their loves and ideals.

The necessity is manifest in the way in which these successfully internalized values will express themselves in the agent's actions and conception of himself as an agent. They will shape his will and thus articulate the limits of his will. They will provide the boundaries against which he will find himself bound and are the sorts of values that underpin the kind of volitional necessities that are at the heart of the practical success of our restrictions against crimes of all sorts. I will argue further in the next chapter that it is these kinds of necessities, informed and generated by volitional structures reflecting successful socialization, that are truly at the heart of most of our daily avoidance of evil.

It is absolutely vital that we recognize that this is not a mechanical or deterministic process. I am not claiming that a developing self is merely an imprint of the world in which it finds itself. It is fundamentally necessary that the developing will itself be fully implicated in this process through its willingness to take responsibility for values which it did not ultimately choose. Unless the agent is subjectively identified with his society's values they will seem to him external impositions rather than the proper expression of his own nature as a moral creature.

This subjective identification, this *taking responsibility*, is achieved precisely by the agent's falling in love with parts of the world. Again, little of this process will be conscious and indeed much of it cannot be. For it is precisely through the integration of a love for parts of the external world (e.g., parents) that the developing will comes into being. The consciousness of falling in love—and of subjecting that growing love to critical evaluation—is only available to agents who have achieved a fundamental level of maturity, and even then it is far from transparent to the person undergoing it.

The content of the developing agent's love will be complex and mixed. As I have suggested his love for external values (be they moral, aesthetic, religious, etc.) will often (though not always) be mixed with his love for his parents. I doubt most agents will have a clear conception of the origin and precise structure of their own loves, nor do I think it is particularly important that they do. For it is in this complicated and rather messy commingling of loves that a unique agency and personal identity takes form.

Ultimately, the complexity of value-laden external world will be reflected in the agent's will insofar as he comes to identify with and take responsibility for those values, making them his own. This may take the form

of falling in love with a vision of a particular way of life (e.g., a religious or moral vision) or the intricacies of some particular social role (e.g., as a highly-trained craftsman), but clearly it is will be different for almost everyone. Nonetheless, there is a kind of minimum adoption of external values that makes it possible even to have an agency capable of valuing—i.e., capable of expressing value through its internal structure. The content of that minimum is fixed to some degree by human nature (and will probably include a kind of mere morality—against murder, stealing, etc.), but I suspect it will also vary both between and within different societies and moral communities, as differing levels of volitional development will be necessary to gain purchase on different societies' diverse social and moral public norms. The more complicated and richer a society's public norms are, the more complicated a fully mature agent's will within that society would have to be.

This kind of falling in love with external values is in some sense pure, because it is only apparent to the agent what has happen after it has happened. Nonetheless, if it occurs as it should the agent will be subjectively identified with his loves and recognize their objective value. However, achieving this sort of self-awareness is precisely the challenge that philosophy has always faced, and to which, it might be supposed, the vast majority of human beings (including philosophers—this one not excepted) never attain.

My argument has thus far aimed at one thing: to show how an understanding of what is required in order for a moral will to develop in a child validates Frankfurt's notion that true autonomy requires necessity. Further I have claimed that this (volitional) necessity is a consequence of, and expresses, the loves the child gains through his childhood which he does not

explicitly choose. I suspect that many will be unimpressed; they will observe that even if my developmental story is granted this does nothing to show that one's childish commitments should continue to hold sway over the will of an adult. In fact, might it not be observed, does not the next stage in maturation—full adulthood—consist precisely in exposing the commitments one has from childhood to a critical evaluation which almost inevitably results in the rejection of some of them?

It is of course true that one's childhood commitments ought to be examined critically to some degree. At the same time, however, it is also a truism that one cannot take up a critical attitude to everything. Nonetheless, the important point for my purposes is that this examination must necessarily take place against the background of certain fixed commitments, to the truth, for instance, but also necessarily in light of the other sorts of loves which constitute mature human beings. As I argued in the last chapter, following Frankfurt, there is a degree to which certain commitments must be free from a certain kind of doubt in order to be the very sort of commitments they are. As he noted, for many deep commitments, including the commitment to a moral ideal of a certain kind of world or certain kind of life, we are ultimately faced with a question more of confidence than of truth.

Thus, this sort of doubt is best viewed not as an epistemic problem, to be solved with greater knowledge. Rather, the "solution" or, better yet, the proper response is a volitional response, a response constituted by a firmer commitment, a more expressive, wholehearted love. Doubt undermines agency precisely by convincing the agent that he must have more grounds for acting than just his love; but this in itself is a failure of love and a failure that

can only be redeemed through a redoubling of love, with an increasing awareness of both the value of that love in itself and of its object.

This does not mean that the morally mature agent is completely unreflective with regards to his deepest commitments. In fact, he cannot be fully unreflective concerning them and still take responsibility for them in the right sort of way. Being deeply committed to something does not preclude reflecting on it, and indeed to fully take responsibility for it will require the agent to have some knowledge of his volitional structure and how his deep commitments fit together. The *mature* moral agent cannot (or at least should not) have a will marked by deep inconsistencies in what he loves, for then his will is perpetually divided and his agency is weakened. It is only through reflection on his commitments that the agent can achieve this kind of maturity.

My view does not even imply that the mature agent needs to be free from doubt concerning his deep commitments. Rather the point is that he must move past his doubts and not be paralyzed by them. Oftentimes moral courage is expressed precisely in moving past doubts (sometimes, but not always, rationalizations of acting contrary to what one suspects is the right thing to do). That is why it becomes more an issue of confidence than truth. Of course committing oneself wholeheartedly does not guarantee one freedom from error (and one can wholeheartedly do evil), but some sort of deep commitments is necessary to even begin to express agency. Deep commitment to (i.e., loving) the truth seems to me such a commitment. Ultimately, I suspect there are other such necessary deep commitments involving at minimum the sort of mere morality I suggested above (though there will presumably be some variation between moral communities as to what is required to be full participants within them).

If the general outline of what I have been arguing is correct, then there are very real limits on what can count as a good-enough world to facilitate the development of morally mature human beings. Thus to the extent that we would cultivate beings like ourselves, and more importantly beings possessing the sorts of values to which I have been addressing myself, we simply cannot help but maintain a world where our most central religious, moral, aesthetic, etc. values are written into the fabric of social roles and discourse. These sorts of moral values, even though they find their best expression internalized in the wills of agents, must be publicly expressed and manifested else they will never become so internalized.

The obligation to provide for a good enough world becomes incumbent on everyone who has a stake in the continuance of society. Most obviously parents must publicly commit themselves to certain ideals because the beginnings of the child's moral consciousness is intertwined with his love for his parents. Without the parents' public commitment to these sorts of ideals it would be extremely difficult for their child to come to love them. (It is not impossible, of course, because children can learn from others besides their parents. However, it is vital for the child to learn to love the ideal at the knee, so to speak, of someone whom he loves.)

However, these obligations are clearly not limited to parents alone. These are society's values, they are the public religious, moral, aesthetic, etc. values that give content and shape to a society, and all who share them have an interest in their propagation to new generations. Thus, in virtue of their public nature it is intrinsic to the possession of those values that he who loves their object must be committed to their preservation within society. To possess moral maturity is to possess, volitionally, these values. However, if one

possesses those values then one must express them, otherwise they are vacuous. Thus, the public manifestation of those values is an ineluctable result of having them in the right sort of way.

It is precisely because the child's will develops in reaction to and in relation with the external world that the inhabitants of that world are required to display and be publicly committed to ideals. In other worlds, the structure of the world which is "metabolized" by the child and becomes the pattern for the structure of his own will is determined by the publicly available actions, attitudes, and beliefs of society at large. Thus to the extent that the world is to contain this necessary structure of public value, the members of society committed to those values will find that the natural expression of those values consists in their public display. This is more or less just what it is to have a genuine moral community in which individuals share and express social values through their manifestation of precisely those values as embodied in the volitional commitments which inform their activities.

Thus a commitment to values has internal to it a commitment to their public display and expression. The practical upshot of this is that a commitment to the possibility of value necessarily will take the form of active loyalty within a society to the persons, institutions, and ideals wherein values are embodied. Thus, I claim, to have (moral) values entails caring about value itself and thus being committed to the preservation of the social structures that make its inculcation in the young possible. Therefore, public loyalty is an important consequence of moral maturity in the life of a shared moral community.

A natural question arises concerning those who reject a society's particular values. Is such a person incapable of achieving moral maturity? I

think the answer to that question is necessarily complicated and dependent to some important degree on the particularities of individual cases. However, as I have suggested above I do think there are certain values a commitment to which are necessary to have any kind of morally mature agency—commitments at least to the truth and some sort of mere morality. Further, I think that in richer and more complicated societies more commitments will be required to fully participate within them.

This is not a trivial claim. There is I think a deep antinomian and even nihilist tendency in human nature. That tendency can take stronger and weaker forms, from the violent anarchism of the 19th century (dramatized, for instance, by Dostoyevsky in *The Demons*) to the kind of malaise and moral complacency that has seemed to characterize "decadent" societies like ancient Babylon and Rome. The processes of moral maturation that I have been discussing require public loyalty to shared public norms; without them moral agents will tend to be compromised from the beginning and the negative (egoistic) tendencies of human nature will find expression in the underdeveloped wills of those agents.

This does not mean, I take it, that an agent must be fully loyal to every particular facet of his society or that he cannot agitate for changes. All human societies are imperfect in some way and indeed always will be. Reflection on the requirements of justice will naturally lead to the recognition that certain changes are desirable. It is fully consistent with my view that a morally mature agent will come to recognize this and become a meliorating influence in his society. Nonetheless, the mature agent will favor precisely that—melioration—not destructive recreation. As such, some sort of public loyalty to the defining values of that society will be required even of the agent who

wants to change it (e.g., values like a commitment to the truth and some sort of mere morality). After all, the love of such an agent for justice is itself predicated on having a will that can be constituted by love, and if my argument is correct, that requires maturation within a moral community which is able to provide a structure in which authentic moral development is possible.

As I indicated above, confidence in and loyalty to the public norms and values of one's society will not insulate an agent from (moral) error. However, if I am right that human nature sets definite limits on what can be a "good-enough" society for moral maturation, then it will be the case that the morally mature agent will already implicitly have the resources for self-correction. For if the commitment to (at least) the truth and a certain level of mere morality is required for even possessing a human-shaped will then reflection on gross injustice (for instance) will necessitate action in accordance with that love for the truth and morality. Because the ground of the will is love there is a kind of self correction built into loyalty insofar as the loyal agent will be committed to the conditions that make love possible—i.e., a good enough world.

Royce suggested that one's loyalty must never destroy another's loyalty, that underneath all was a loyalty to loyalty. I think the insight Royce has incompletely captured is an implicit recognition that human nature and the public structure necessary for "good-enough" moral development have within themselves the seeds for self-correction. The loyal agent is not indefeasibly committed either to every single contingent public norm of his society or to the perfect stasis of that society within some imagined perfect past. What he is committed to is the maintenance of the kind of society that

does properly foster the development of agents committed to the truth and the natural good. Thus, even the reformer must remain loyal to some degree.

This may not apply, however, to societies that are not "good-enough" in the relevant sense. A sufficiently unjust society may forfeit the loyalty of its members just insofar as it is unjust. A commitment to maintaining deep injustice will conflict with the fundamental volitionally-constitutive commitments to truth and mere morality. An agent characterized by such inconsistent commitments will have a conflicted will, and his agency itself will be seriously compromised. In this case, paradoxically, loyalty may very well entail disloyalty to intrinsically disordered public norms. If the injustice is radical enough loyalty may even entail the destruction of that society.

Loyalty to deep injustice would be disloyalty to the ideal of the sorts of public norms necessary to develop truly morally mature agents. Therefore, loyalty to such norms would be destructive of the deep loyalty that is, I have claimed, foundational for human society itself. The agent in a society with such intrinsically disordered public norms then has a reason (from loyalty to the truth and mere morality) to display his opposition to those disordered norms. This, however, does not undermine the claims I have made for loyalty, quite simply because it is ultimately private loyalty to some higher end that motivates this public disloyalty.

However, I think these cases are unusual. A fully mature agent who embarks on such a revolutionary course should be very reticent and never forget how extreme social upheaval can have unforeseen consequences that radically undermine the possibility of realizing the very justice and truth that his deep commitments drive him to attempt to achieve. I think there ought to

be a kind of prejudice in favor of loyalty to one's society as it exists, but a prejudice that can ultimately be superceded in extreme cases. [199]

In a recent essay on Frankfurt and "love's authority" Lear makes the following observation: "Frankfurt's central idea about love's authority is this: The lover, in acting against the dictates of his love, ultimately betrays himself" (Buss and Overton [2002], 279). When we combine this with Frankfurt's notion that our moral activity itself proceeds from "love for a certain kind of world or a certain kind of life" I think the connection between mature active loyalty and the developmental story I have been telling becomes clearer.

If our moral activity comes out of our volitionally constitutive commitment to a certain kind of moral ideal, and that requires the kind of dialectic between developing will and world that I have described above, I think it becomes plausible to understand why external loyalty is ultimately a loyalty to oneself. Because the internal structure of the will of a mature (moral) individual is a reflection of the structure of the world in which he attained maturity, an internal requirement of consistency—loyalty to oneself—will be reflected in the reverse direction as loyalty to the moral superstructure of the external world. Because the inside reflects the outside, so to speak, loyalty inward will be externally manifested in loyalty outwards. Thus, private loyalty to oneself becomes public loyalty.

Loyalty to persons, institutions, ideals, etc. will proceed necessarily from the possession of a moral consciousness. Because what it is to be

[199] The legal aphorism that hard cases make bad law is surely relevant here. Although I want to leave open the possibility of these extreme cases (e.g., the anti-bellum South or Nazi Germany), I do not think we focus our attention chiefly (or even substantially) on such cases. I think it is much more important for us to get a feel for how loyalty would work within a "good-enough" world.

possessed of a moral consciousness has internal to it a concern with its own preservation, a recognition of the nature of how one's internal values are necessarily connected with public values will lead to a recognition of an internal demand to conserve the public structure of value.

Moral maturity consists in having a will that embodies these values. Having such values entails valuing them as such and valuing them as such entails caring for their propagation (because they are intrinsically social, public values). Finally, caring for their propagation (because that involves the complex dialectic of love and loss described above) requires public loyalty to them and the moral community they inform and embody.

On the picture I am painting, ideals (moral and otherwise) exist both internally in agents and externally in the world. It is precisely because these ideals have internal to them their own preservation and propagation that they generate an internal demand for external loyalty. As we come to recognize this through an examination of the developmental process, I think we can come to see how important the stability of public values are for the inculcation of value in future generations. The very structure of human moral development requires it.

In light of these considerations, I hope it is now becoming slightly clearer how the notion of a life with "substance," with which I began Part II, will have particular and inescapable requirements for how the will of a morally mature human agent must be structured, and thus how the external world must be structured to accommodate this fact. On the other hand, however, these structural requirements are quite broad and do not prescribe all the particular cultural formations of our or any other culture. What they do

describe is how any human culture must possess certain limiting public moralities and commitments which will serve as school of virtue in its young.

Although I do not believe human flourishing requires Western civilization as we know it, I do believe that, because of the way in which the development of a (moral) consciousness centrally involves learning to love a certain kind of (moral) ideal, any given culture, including our own, has an internal necessity towards its own preservation and propagation, embodied in the volitional structure of its members. Thus, to have fully a life of "substance" in our culture and civilization requires a commitment and loyalty to it, precisely because the substance of such a life is partially embodied by the same complicated public/private values that embody that civilization and culture.

Thus, I return to a particularism about values. This particularism is not a relativism, however. Human nature will prescribe inescapable limits to what sort of social structures will make possible the development of the kind of wills capable of moral striving. Within those limits, however, much diversity is possible and indeed desirable, precisely because this diversity will point the way towards understanding what the limits of human nature are and thus who we are as human.

However, within a given sophisticated culture like ours the appreciation and preservation of the religious, moral, aesthetic, etc. ideals that inform that culture will require the education of souls—i.e., the dialectic of love between developing soul and external world that I have tried to describe in this chapter. As such, the more developed the values a society embodies, the more important it becomes for the maintenance of those values that there is a level of public stability and confidence in them. Without that sort of

stability it becomes increasingly difficult to maintain the vitality of that society and the values it embodies. Without a straightforward love of these sorts of values, moral agents will find their very agency undermined and the possibilities for true moral striving, for lives of genuine worth, increasingly lost to relativism, apathy, and despair.[200]

In this chapter, I have tried to describe the highly complicated interrelationship of intra-volitional structure and the structure of the external moral and social world by discussing how an understanding of the nature of the development of the will prescribes certain limits to the world. I hope this has suggested that the demand for the justification of our loves and loyalties is aimed in the wrong direction. The justification goes no further than that such demands are the inescapable consequence of having a will shaped by love.

In the next chapter, I will try to make good on my promise at the end of the last chapter to integrate the insights of the hierarchical model of the will with the developmental concerns that have been my focus in this chapter. I will discuss certain difficulties in the Frankfurtian account that only an understanding of the deep inter-relationship of will and world can resolve. Finally, I will tie together a variety of loose ends and attempt to show why my concerns in these last two chapters point directly to a more fruitful understanding of the genuine nature of loyalty.

---

[200] Again, this does not entail a mindless conservatism of whatever happens to be the case. It does, however, entail a thoughtful conservatism, in which the values and interests of both the past and future members of that society (to borrow a thought from Edmund Burke) are important considerations and worthy of protection. Ultimately, however, the demands of a commitment to truth and mere morality will naturally guide the morally mature agent in pursuing the melioration of social imperfections.

CHAPTER VII

A Developmental Argument for Loyalty

It is now time to return our discussion to the questions at the heart of this dissertation. In the last two chapters I have introduced several new tools; now I want to utilize those tools in constructing a positive argument for my conception of loyalty. Along the way I will address several remaining objections and articulate in greater detail the importance of the developmental account in understanding the structure of a human shaped will of the sort we have reason to value.

In Chapter V I discussed Frankfurt's theory of the will. I believe that Frankfurt's theory offers the best hope for understanding what is really at stake in the deepest sorts of human loyalties. These are the loyalties in which the implicit connection between the agent's acts and identity are brought into relief. I argued there that the deepest loyalties in human life are those which involve values so deeply embedded in the agent's will that the actions which express those loyalties are essentially acts of *self expression*.

The focus of Chapter VI was on how such selves come to be. Taking up and expanding on Lear's discussion of the development of the nascent psyche, I argued that the moral psyche—the ground of moral agency—comes about in an analogous way, through a dialectic of love between the soul of the

developing agent and the loving and lovable world he inhabits. This fact, the deep intertwining between the individual's soul and the world in which he came to be, points us towards seeing past a potential misconception. For if the moral self and the world in which it develops remain in this sort of dialectic, then the self-expression of loyalty is not merely a kind of subjectivism. There is a specific non-self-regarding content to these deep loyalties, a content which embraces the moral structure of that loving and lovable world with which the individual soul is in dialectic.

By 'dialectic' here I mean just to emphasize the way in which the values in question, as I argued in the last chapter, are both public and private, societal and intra-volitional. The dialectic in question occurs between the individual's will and the external value-laden world whereby the individual takes up or "metabolizes" social values through his coming to care for them. In so doing he "writes" them into his soul insofar as they become the same sort of volitionally constitutive commitments discussed earlier.

It is about this connection, between loyalty as self-expression and its role in mediating between the internal structure of the will and the external structure of the world, that I want to be particularly clear. I think this is where my approach has the most to offer, for this is where it should be the most apparent just why gaining an understanding of the nature of volitional development is so important.

This connection may not be entirely clear on its face, precisely because these two aspects of my argument are designed to meet different issues. My focus on the constitution of the will is aimed at the phenomenological, subjective side of our deepest loyalties. What it shows is that the *felt* importance of these loyalties is not merely superficial but rather reflects

something important about the role that loyalties actually play in the volitional life of the moral agent.

We tend, rightly, to regard an emphasis on the felt quality of certain experiences with suspicion. Indeed, this sort of complete absorption in a felt experience is commonly seen as a sign of immaturity. One need only think of teenagers' common complaint that others just do not understand them, because those others do not know what it is like to *feel* as they do. This sort of solipsistic "argument" tends to be immune from criticism; of course one cannot precisely feel someone else's emotions. Nonetheless, part of moral maturity, indeed maturity of any sort, involves getting beyond being overcome by one's feelings.

The fact of a feeling, with its attendant subjective felt significance, is not enough to justify its importance, all things considered. We commonly meet with individuals absorbed in their feelings who strike us as fundamentally shallow and unserious. My purpose here is not merely to pick on such unfortunate souls but to make a simple (and obvious) point: the subjective feeling of importance is not enough to justify a loyalty. Therefore, if my argument for the deepest loyalties of human life turned merely on their felt importance, then it would have little more weight than the plaintive cry of the teenager's "you just don't get it!" For us to take them seriously loyalties must not only have great felt importance, they must embrace serious values important beyond their subjective appearance.

I would note in passing that the relationship between felt importance and volitional constitution goes both ways. It is a feature of Frankfurt's model that, as a formal model, it does not really place limits on what sorts of commitments can be volitionally constitutive. In the case of the teenager it

may very well be that his commitment to the fortunes of his favorite band "rises" to the level of a volitional necessity in terms of its felt importance. What this reveals, however, is not the importance of that band but the poverty of the fan's soul. His will actually may be constituted by such frivolous commitments, but obviously that does not, by itself, justify such commitments. Indeed, it only makes us realize just how uninteresting he is. In other words, as we ought to expect, the degree to which the agent embraces the frivolous as subjectively important to his life only reflects the degree to which the will and self at the ground of his agency is fundamentally immature, impoverished, and uninteresting.[1]

This seems to point to a deficiency in Frankfurt's model insofar as it is a model of rational free agency. The single most important feature of that model is Frankfurt's claim that free agency consists only in acting in accord with one's higher-order, volitionally constitutive commitments. It does not really seem to matter what those commitments are. In other words, Frankfurt's is a formal model, in which rational free agency is a product of having a certain volitional structure independent of its content.

I think the deficiencies of this formalism can be brought out in two ways. I have already adverted to the first in the case of the frivolous commitments of the teenager. The second lies in the vulnerability of the will so conceived to deliberate external subversion.

Remember that Frankfurt's original goal was to give an account of freedom of the will. He wished to understand that freedom as consisting in

---

[1] I am not suggesting that a commitment to music, or even to some particular performers, need be necessarily immature in this way. Nonetheless, I think most of us have met someone (or indeed been) like the teenager in my example. Any sort of overwrought commitment to some frivolous end may be substituted to see the point. Certain sorts of Gothic and Romantic literature come to mind in this regard (the sort mocked by Jane Austen, for instance).

being guided by higher-order commitments to which the agent is reflectively and wholeheartedly committed. Freedom (partially) consists in not being compelled by (first-order) desires over which one does not exert reflective control. However, Frankfurt never truly addresses where these higher-order commitments come from and surely this poses a *prima facie* problem.

Although it seems that Frankfurt is right to claim that freedom partially consists in the absence of internal compulsion—possessing self-control over one's first-order desires—freedom just as clearly also requires the absence of external compulsion. However, it is not clear that Frankfurt's model conclusively rules out a case in which an agent's higher-order commitments may be forced on or implanted in him. Why might not an agent be reflectively committed to higher-order desires which were somehow externally introduced into his will? If that were the case, it seems unlikely that we would consider him free even though the shape of his will would meet Frankfurt's criteria.

It might be objected that this is impossible, that the nature of reflective commitment precludes this. I do not see why this should necessarily be the case. It might be impossible to introduce a single higher-order commitment in conflict with others that the agent already has. However, why would it be impossible to completely remake a given agent's set of higher-order commitments through some sort of extreme psychological treatment (or perhaps one day through some intervention in the physiology of his brain)? So long as this set itself were not internally incoherent the agent would be able to subject these higher-order commitments to rational reflection and deliberation. He could reflect on those commitments and would not be aware of any disturbing conflicts. Indeed he could come to acquire other

commitments congruent to (or at least consistent with) them, and he would be utterly unaware of any abrogation of his freedom.

In my example I am supposing that the agent's entire set of higher-order commitments (or a large fraction thereof) are directly introduced into his will by an external agent. This may seem so far-fetched as to fail to present a genuine objection. However, extreme psychological manipulation is certainly possible. We need only consider a Stockholm Syndrome case like Patti Hearst. In that case, through physical and mental intimidation a kidnapping victim ultimately came to sympathize and identify with her kidnappers. She began to take on the kidnappers' values and even to participate in their crimes. By whatever particular modality this came about, it seems right to regard this as an external compulsion. It is very implausible to consider her reaction truly autonomous in the way it would need to be to underwrite an authentic expression of the victim's will.

I think it is important to note that there is nothing in the content of these particular commitments that make them incompatible with autonomy. In other words, it is reasonable to suppose that the kidnappers themselves possessed sufficient autonomy to be morally culpable for their actions.[2] It does not seem to be the case that one cannot autonomously hold such commitments.

We should keep in mind that Frankfurt's model is in some important ways very limited. He purports to give a structural account of the will that

[2] This potentially leads to a tangled thicket of issues concerning the relationship of autonomy and moral culpability. For my purposes here I am assuming that one can autonomously choose to do evil, and thus be morally culpable for it. Now it may be the case that someone choosing evil is *ipso facto* incompletely autonomous. Nonetheless, even if there are relevant degrees of autonomy, it is seems to be the case that there is some level of autonomy necessary for moral responsibility which evil-doers meet.

explains volitional autonomy. The problem is that the bare structure of a free will as he develops it is compatible with that will having, even at the higher-order, commitments introduced by an external interest. It is precisely because the model is limited to a *structural* description of volitional freedom that it is vulnerable to this kind of manipulation.

If there is nothing about the particular content of these commitments that distinguishes their autonomous instantiation in the kidnapper and their non-autonomous instantiation in the victim then there must be some other grounds for our intuition that in cases like these the kidnappers are acting freely and the victim is not. My suggestion is that we need to look at how these commitments come about in a given agent's will. It is the nature of the process whereby they are acquired that distinguishes the two cases.

In the case of the kidnappers I think we would want to say that their criminality was an authentic expression of what they cared about. If we were able to examine the wills of the people in question I suppose that we would find that their actions are consistent with ends that they themselves identify with, and that those actions reflect (dis-) values that they autonomously hold. In other words, their criminality would presumably be consistent with their past lives and reflect the structure of their wills.[3]

In the victim's case, however, her new commitments are radically discontinuous (we may suppose) with the constitution of her will prior to her abduction. They are radically discontinuous with what we might suppose she most cared about previously. Accordingly, it seems unlikely that she could

[3] Actually, I want to prescind from the particulars of this case, because they are not so important for my purposes; suffice it to say that whatever was the case with the actual kidnappers of Patti Hearst, it is perfectly possible to imagine kidnappers with the sort of volitional constitution I am here describing.

have autonomously developed these new commitments. As I have argued earlier, because one's volitional commitments are the very ground from which one's agency proceeds, it is impossible to choose to completely remake one's will. Truly autonomous choices must proceed from one's central commitments. Of course one's central commitments can change and evolve over time, but a complete remaking of them seems impossible.

I wish to prescind from the exact details of the Hearst case, because what matters for seeing my point against Frankfurt's formalism is not the state of her particular will. What matters is that we can see that his formalism cannot adequately explain a case of this sort whereby the agent's will seems to be so remade that it is highly discontinuous with her previous (autonomous) will. We need some other account of why there seems to be a radical violation of autonomy in this sort of case; why the mental, emotional, and physical pressure that eventuates in such a change is inherently coercive, even if the agent later seems to be committed to values of her tormenters.

My developmental model focuses our attention on how values are legitimately taken up into the will through the dialectic of love and loyalty I have described. What matters there is that the dialectic is precisely a developmental process, whereby the nascent agent's very will is in the process of forming in reaction to and in relation with the public values of the good-enough loving world in which he is maturing. In the coercive case we have commitments being imposed through a kind of violence that threatens and destroys the agent's pre-existing commitments. In the developmental case, the development, to be successful, must be inherently loving (broadly speaking) and thus is not intrinsically destructive or violent in the same way. The developing agent does not have pre-existing higher-order commitments to be

destroyed but is in the very process of acquiring those commitments for the first time.

What all of this suggests is that Frankfurt's account is incomplete insofar as we hope to have a full understanding of autonomous agency. As a strictly formal model it is consistent with an obvious case of coercion. Thus, it seems to me the question of how an agent develops his higher-order commitments is vitally important. An agent's will is free not merely if his higher-order commitments are internally coherent and reflectively his own, but also insofar as they are not implanted within his will by external forces. This then is why I place so much weight on the developmental account of moral psychology. It is meant to describe the natural generation of *freely acquired* higher-order commitments that may still be unchosen.

These commitments are freely acquired insofar as the agent himself comes to take responsibility for them—claiming them for his own agency. They are nonetheless unchosen insofar as this is a developmental process such that the nascent agency in question is in no position to choose the commitments, as it is literally coming to be through them. To speak of such an agency *choosing* is to mistake what it is that choosing consists in. For, as I discussed previously, such a nascent agency has no stable ground from which to make choices, no stable core of volitionally constitutive commitments from which a genuine choice can proceed as reflective of a self. This is precisely because that nascent agent is undergoing the dialectical process whereby such commitments are first instantiated. The etiology my account describes shows how genuine higher-order commitments can grow out of an unchosen ground yet nonetheless be the foundation for a real volitional freedom.

In light of these two distinct but related objections to the bare formalism of the Frankfurt view, I hope the value of the developmental model will become more apparent. For developmental considerations offer a way out of some of the problems while nonetheless preserving the features of the Frankfurt model that are central to my concerns about loyalty (viz., the determination of the will according to volitional necessities). The developmental model addresses the first concern (the problem of frivolous volitional commitments) by showing how a mature moral agency is the product of a dialectic with a "good enough" moral world. Loyalty occupies a central place in this dialectic for it is the means by which the individual acquires public values and makes them his own.

Now this does not mean that a good-enough world will completely supply the content of the mature agent's will. I do think it is perfectly possible for mature agents to have frivolous commitments at the periphery of their wills. So it is possible to have a morally mature agent committed to something that may seem to have little objective value, such as the fortunes of a favorite rock band. However, these cannot be the central commitments of his will insofar as he is able to participate in his moral community.

In other words, the good enough world does not prescribe the full structure of the will of the mature agent (and indeed we should not want it to, for otherwise we would be left with moral automatons more than moral agents). Rather, the good enough world will prescribe the central commitments that make socialization possible—e.g., commitment to the truth (broadly speaking) and some sort of mere morality. These, in turn, make our other commitments (trivial or serious) possible, for they underwrite the possibility of social engagement and discourse. To meet the objection to

Frankfurt, the agent's response to the good enough world does not need to *rule out* trivial content; it just needs to *rule in* these sort of grounding commitments that make the moral life possible.

When we realize that the mature will can only come to be within such a dialectic we can begin to see that substantial limits are placed on the content of such a will. These limits force the developing moral agent to transcend the limits of his narcissism, precisely because the values which they represent are public values woven into the fabric of a shared social life in which the discourse and practice of morality is intrinsic. In other words, the price of participation within this shared moral life is precisely the development of a will in which the same values which inform that social life become internal structural components of that will.

The shared social life of public morality is only possible because the agents which inhabit it have successfully internalized its basic constitutional forms. I think the artificiality of so much modern moral philosophy brings this into stark contrast. When we turn our attention to moral choices as we actually make them we ought quickly to realize just how far our choices are already limited. So many bare possibilities that might be thought available in the abstract consideration of the seminar room are simply not really possible as real world choices.

What I mean by this will be clear after a moment's consideration. Let us begin by examining why morally mature agents in stable moral communities do not exemplify so many of the little vices which are often rampant in areas where public morality has broken down. Consider something as simple as littering. No doubt various arguments could be constructed to show an agent genuinely undecided on the question that he ought not litter—that the

ultimate disutility, even to him personally, of a cluttered environment outweigh the present utility to him of dropping the sticky candy bar wrapper. However, it should be obvious that having been convinced of such an argument or similar ones is not the reason most people do not litter.

For a well brought up—i.e., morally mature—moral agent it simply will not enter into his mind that one of his options for disposing of the candy bar wrapper is to drop it on the ground. It is clearly not the case that at some point in his life he reviewed the arguments for and against and formed a standing intention not to litter. It is much more basic than that: it is that certain actions simply are not possible for him, because they are inconsistent with social norms he has so internalized that doing such things would be deeply alien to his character.[4]

I cannot help but think that moral philosophers do not give nearly enough consideration to this phenomenon in their discussions of moral motivation. For I believe that the reason I do not murder people who annoy me is often very similar to the reason I do not litter—neither option ever really enters into my mind. This is of course not to claim that murder is of equal moral import to littering. Rather, the point is that in the practice of moral life, lived moral existence in the shared social world, what regulates our moral existence is rarely, if ever, a conscious, reflective commitment to the good (even when we have such a commitment). What regulates the activity of life is a standing disposition to act in certain ways because the scope of possible

[4] Obviously, extreme circumstances might change this. If he sees a child fall into a pond, he will hopefully forget his anti-littering instincts long enough to jump in after him, even if that means the candy bar wrapper hits the ground.

action is already circumscribed by the nature of our wills as formed within society.[5]

It is this fact, that the scope of moral decision making is already highly circumscribed by volitionally internalized social norms, that is of signal import in considering the relationship of the structure of the will to the developmental model of moral psychology. For the developmental model articulates how this comes about, how an agent's will takes on these internal constraints in the process of normal development. Insofar as that will takes on the structure of public norms through development in a good-enough world, those norms will then condition and inform both his dominant action guiding dispositions and the scope of possible deliberation. It is these public norms that will define the scope of readily imaginable possible actions in determining just what will even enter the mind of the agent. Further, to the extent that we are able to understand the organizing features of the social fabric we will understand what sorts of values need to be volitionally internalized in order to ensure moral maturity.

Without a consideration of these developmental aspects no model of the will can be faithful to the nature of human wills as we find them abroad in the world. In particular, no formal model that is completely agnostic about the content of the will can ever hope to be a true representation of the sorts of

---

[5] Bernard Williams alludes to this idea that certain things never enter into one's mind as a live option under the title "deliberative silence." In *Ethics and the Limits of Philosophy* he notes:

> An effective way for actions to be ruled out is that they never come into thought at all, and this is often the best way. One does not feel easy with the man who in the course of a discussion of how to deal with political or business rivals says, "Of course, we could have them killed, but we should lay that aside from the beginning." It should never have come into his hands to be laid aside. It is characteristic of morality that it tends to overlook the possibility that some concerns are best embodied in this way, in deliberative silence. (Williams [1985], 185).

wills that we actually do value and thus wish to understand. For what we ought to be most interested in are the wills of agents who are morally mature; that is, of agents who successfully inhabit and engage the social world we have reason to value (namely, our own).

What explicit consideration of the developmental model shows us is that individual agents cannot be understood independent of the environment of moral maturation. Indeed, individual moral agents are to a large degree incomprehensible without such social context. Most modern moral theories tend to take the agent as a given (though what is given is variously understood). Kant sees the agent as a will, an instantiation of pure reason that is a cause within the "noumenal" world. The utilitarian sees him as a bundle of pleasure-seeking desires. Rousseau sees him as driven by *amour propre* and the desire for recognition. Though the details of the theories vary widely, they each tend to base their reflections on motivation and moral duty on fully formed moral agencies.[6]

Indeed contemporary moral theories often classify the sorts of issues I am here discussing as pre-moral. They are the given on which the moral philosopher works and to which he applies his reflective moral theory. As such these sorts of issues are generally considered not worth discussing. I think this is a mistake because it treats these developmental issues as if they were merely brute, which the moral theorist can then refashion for his own purposes. Such an attitude ignores the fact that the processes of development

[6] Rousseau is something of an exception to this, of course. In *Emile* he explicitly considers the problems attending a moral education, and in the *Second Discourse* he has a sort of developmental model for the species. Nonetheless, I do think it fair to say he does not really concern himself with a full-fledged developmental moral psychology.

will determine, to a large and perhaps decisive extent, the real scope and nature of moral activity in the world of lived experience.

What the developmental model allows us to see is precisely that the result of moral maturation within a loving and lovable social world has inescapable consequences for the shape of the wills of the agents who develop in it. This shape is determined by the way in which development within a social world results in the internalization of that world's values as volitionally constitutive. In other words, when an agent reaches moral maturity his will has already been greatly formed; he possesses a variety of volitionally constitutive commitments which reflect the nature of the social environment in which he matured and to which he will then naturally be loyal.

As I have repeatedly emphasized, loyalty is both the means for, and the product of, this moral maturation process. It is a means insofar as the nascent moral psyche itself comes into being by internalizing the external social values through loyalty to them (perhaps beginning with the most basic kind of loyalty—identification—but moving beyond any sort of mechanical process to a volitional one). It is an important consequence of this process that a will formed in this way will naturally "express" itself through the agent's public affirmation of the social values internalized during development. Ultimately, the most reflective agents (which are not necessarily the best, by any means)[7] may progress towards reflectively embracing those values through something like a will to believe. In other words, loyalty lies at the very heart of the

[7] I tend to agree with Williams that the relationship between reflection and moral knowledge can be negative, that reflection can destroy knowledge (see note 9 below). Whether that is a bad thing or not is another matter. However, irrespective to the question of that relationship, it is often the case that reflection is the enemy of confidence and conviction, than an over keen ability to reflectively examine all sides of an ethical issue may eventuate in paralysis or a sort of nihilism.

possibility of a social and moral order because in loyalty mature agents are tied together, and through loyalty the young are able to take up their places within that order through the processes of moral maturation.

This does not preclude a kind of tempered rebellion at certain points. Teenage rebellion is a cliché, but of course reflects an underlying psychological and moral reality. To accept fully the social values into the will in a manner that will underwrite genuine rational agency many people seem to need to push them away for a while, to find their "own" way. In a stable and self-perpetuating society most of these people come back to the very values they apparently rejected, but they do so perhaps with the advantage of a space that allows them to claim them for their own—to take responsibility for them—by seeing them from the "outside." Before they gained that "space" through this sort tempered rebellion they may have felt themselves victims of a "cookie-cutter" morality; with the benefits of that space they can come to identify with those values without feeling compelled.

This sort of tempered rebellion is often important in generating individuality. It is in pushing back the demands that the world makes that the individual comes to express his own unique will and agency. Even the two year-old's emphatic "No!" serves an important purpose in helping him to see himself and his action as independent sources of value. Ultimately, some such independence is at the heart of moral responsibility. Nonetheless, in the end, the communal good, and thus the good of individuals, is only possible to the extent that this tempered rebellion gives way to an integration of the public values within the agent's will.

Of course this need not deny the potential creativity of the agent. Public values do change (for better and for worse) and these changes are often

wrought by individuals demanding greater recognition of their individuality. Thus, nothing in my view requires public values to remain firmly fixed. However, it is important that there remain a kind of core of shared public values which are fixed and which ground the very possibility of shared moral community. These include the kind of values that are necessary to negotiation and compromise, etc. In the long run, I think they will also have to include commitments to the truth and the sort of mere morality I mentioned previously.

The upshot of these simple observations is actually quite important. No social order, however conceived, can long endure, and remain a repository of moral value, without an active and public loyalty expressed by its mature members and inculcated in its children. Therefore, any moral theory that fails to pay consideration to the demands of loyalty in the maintenance of moral community is, at best, myopic, and at worst, positively pernicious.

Even if it is not the intention of the authors of such theories, the effect of the common sort of moral theorizing is quite often detrimental to the maintenance of moral community. The constant search for principles of justification grounded in putatively independent principles of the good (e.g., pure practical rationality or pleasure maximization, etc.) can deeply undermine the processes of moral maturation. Because they force the agent to seek justification only in terms of abstract (generally universally quantified) principles, they radically undervalue the degree in which the real moral life is lived and shaped in the context of volitional commitments unconsciously acquired through the processes of moral maturation. In other words, it is simply not the case, as one might suppose from so much moral philosophy,

that the key to acting morally is the constant justification of one's actions (or even one's character) by reference to putative principles of the good.[8]

There is something intuitively very wrong in the picture that many modern moral theories often seem to paint of moral reflection. The agent whose focus is always on the good alone seems deeply inhuman somehow. I noted in Chapter II Michael Stocker's discussion of the apparent incompatibility of so much of modern moral theory with the plain reality of lived ethical lives that are centrally focused on particular attachments, friendships, and loves that the agent simply does not justify according to external principles of the good.

One of the key advantages of trying to understand the development of moral psychology is that it forces us to recognize this fact. When we are clearly dealing with developing moral psyches in the process of becoming agents we are more readily able recognize the dependency of the individual will on the social context of the loving and lovable world. When we consider just how much the developing psyche must achieve in order to attain the status of a free agent, we are brought to see the degree to which that complexity must arise from some clearly external source, for the resources simply are not there for it to be self-generated. Finally, by taking the process of maturation to be one in which the public social values of the world are connected to the developing psyche through his loyalty to them, we can come to recognize how deeply embedded the moral agent is in the social discourse

[8] I take it that hardly anyone would believe that *every* individual action would have to be so justified, but I do think that most moral philosophers would think that one must have, at least, standing intentions, reflectively justified according to such principles, for regulating moral action. See as well note 3 above concerning Williams' notions of "deliberative silence"

and practice of morality, independent of which there is no moral meaning nor the possibility of a full moral life.

Thus, if the advent of moral agency is only possible through a kind of loyalty to the public values of society, then moral philosophy, whose ultimate value must lie in how well it serves the practice of moral life, cannot legitimately undermine that loyalty. For with the undermining of loyalty comes the undermining, intentional or not, of the very possibility of stable moral community. In other words, what is at stake is not merely loyalty to some given, contingent moral community. What is at stake is the volitional structure—including the internalized public norms—necessary for maintaining and perpetuating any sort of moral community whatsoever, e.g., commitment to at least some sort of bare morality governing social interaction. No actual moral community necessarily commands absolute loyalty just because it exists, but loyalty to some set of shared public norms is necessary for any moral community, including the potential ones imagined by a reformer.

The chief "liabilities" of loyalty as a moral phenomenon—that loyalty is putatively unthinking and merely dog-like, that it is often unjustified according to independent moral principles of the good—are in fact great strengths insofar as the inculcation and maintenance of moral community is concerned. For only insofar as agents are almost instinctively loyal will they uphold the public norms that make possible the survival of the moral community in the face of both internal and external threats *and* the self-perpetuation of that community by inculcation of those public norms in the young through the very developmental processes I have been at pains to describe. Loyalty provides the ground condition for *community* as opposed to

the mere collection of moral agents. For loyalty binds agents together in a public structure that has reasons and interests greater than the mere aggregation of the reasons and interests of its constituents. If it is true that moral agents need a moral community in order to be moral, then it is the case that loyalty to community is necessary for the very possibility of a lived moral life.

This is by no means to imply that the proper course is to abandon moral reflection and moral philosophy. Rather it is a warning to those engaged in moral philosophy that it is a mistake to think that only those volitional commitments which can be justified explicitly in terms of external principles of the good are proper to an ethical life. The very substance of a sustained ethical life, much less sustained moral reflection, is parasitic upon the stability of the moral community and thus upon the stability of the volitional commitments of those who make up that community to uphold it *as a moral community*.[9]

At this point I think it would be well for me to address an objection that readily arises towards my emphasis on development. The objection can be stated simply: exactly why should the modality whereby a moral self comes to be continue to have moral significance for the adult self? In other words, why

[9] In his *Ethics and the Limits of Philosophy* Bernard Williams arrives at a very "un-Socratic" conclusion: "that, in ethics, *reflection can destroy knowledge*" (Williams [1985], 148). He does not therefore argue that we must stop moral reflection, because the loss of ethical knowledge is compensated for in other ways—e.g., knowledge about other areas of human life. Nonetheless, he argues that what is required is *confidence*, and that moral practice can be founded on a confidence that does not depend on epistemological certainty concerning moral *knowledge*. Williams is quite dismissive of moral theory as such, arguing that moral theory, by stressing obligation, militates against other forms of the ethical life, embodied in "practical necessities" of other sorts. Hence, the "limits of philosophy" (particularly moral theory) in articulating the ethical life are quite real.

is the means by which an individual comes to possess a fully responsible moral psyche not simply like Wittgenstein's ladder—to be kicked away once ascended? Indeed, might we not be tempted to remark with St. Paul, "When I was a child, I spake as a child, I understood as a child, I thought as a child: but when I became a man, I put away childish things"? Why then, even if the outline of my developmental story is granted, should the consequences of that developmental story continue to bind the mature agent?

Consider this analogy: one child might learn the rudiments of arithmetic by manipulating pennies under the guidance of his teacher; another might learn the same by counting oranges with her mother. Let us say that each child has mastered (in some sense) the truth of the proposition "2 + 2 = 4"; in that case, is not the modality by which the mastery was achieved irrelevant? Can not each child now, in principle, put away the pennies or the oranges, and apply the algorithm to anything?[10]

That *in principle* is of course at the heart of what is at stake. We tend to think that what matters in the addition case is that there be some principle or algorithm, etc. whose truth is fully independent of any given instantiation of it. Thus, while it may be contingently true that an individual's understanding of the principle is achieved through practice with some particular instantiation of that principle, nonetheless the truth of that understanding is not dependent on that instantiation, but precisely must transcend it to constitute a true principle. The truth of "2 + 2 = 4" lies not in any objects and would be true even if there were no objects.

[10] Scott MacDonald first phrased the objection to me in this way.

Why should not moral principles be the same? If, for the sake of argument, some principle—e.g., "one ought to be loyal to one's family"—is true, what possible difference could it make that I learned this in a family in which my loyalty never had much purchase beyond my parents, while someone else learned it in a family encompassing a myriad of cousins, great aunts, and extended relatives seemingly beyond number? Should not an analysis of such moral principles prescind from the particular and embrace the universal, precisely because that is what makes them principles?

Thus, we might think that just as the truth of the arithmetical proposition is independent of how any given person came to know it, so is the truth and moral force of any moral proposition independent of how any given moral agent came to embrace it. What matters is not *how* the agent came to have a mature moral self, but *that* he did, for it is precisely in virtue of possessing such a self that he is a moral agent, an independent participant in the sphere of moral discourse. In fact, the extent that the agent remains caught up in the particular modality of how he came to understand some moral truth might itself be thought to be the degree to which he has failed to understand it fully.

It is important that we recognize that this objection has force because of two key assumptions, namely that the moral self is an independent source and bearer of value and that morality has purchase on that self through its subsumption under universal moral principles. If it were the case that the individual moral self were not so very independent after all and that morality did not necessarily have moral purchase on selves simply through their subsumption under universalist principles, the force of the objection would drain away.

Of course, the model of the self I have advanced fits this description of the fundamental embedding of the agent in the moral community. In the last chapter, I essentially argued that the values which generate a mature moral self through the developmental process of taking responsibility are public values. Their publicity should be understood as two-fold (which are really two sides of the same coin): first, they must be public for them to be effective in forming the nascent moral self through the dialectic of love. In other words, it is only because they are abroad in the world that the developing self is able to claim them for himself, i.e., able to integrate them into his developing soul by identifying with and then taking responsibility for them (through loyalty to them). Obviously, if they were not public they could not participate in this dialectic, for the developing self would have no access to them. It is precisely in virtue of their publicity that they can be effective objects of loyalty for the developing psyche.

The second sense of their publicity lies in how they manifest themselves once integrated into the developing self—i.e., once the agent has achieved maturity. *If* they have in fact been integrated (it is perfectly possible for him to reject many of them) then they will manifest themselves in his social behavior. For these values to have become integrated into his will through loyalty they must be what the agent most cares about. Accordingly, to the extent that his actions reflect the structure of his will—i.e., to the extent that he most expresses his genuine autonomy—he will publicly manifest these values.

If the developmental account that I have tried to give is on target, then many of the values that inform the will of any given agent will have to be public values of this sort, because that agent's will is itself partially a product of those values (i.e., they have been "metabolized" by the agent in taking

responsibility for them). I have in mind here values of all sorts, from the unambiguously moral (e.g., the prohibition of murder) to the social (e.g., civility in social discourse) to the aesthetic (e.g., an appreciation for opera). What this suggests then is that even as integrated into the soul of the mature agent these values will retain their publicity. Indeed that they are the values that they are—values that inform and maintain communities—requires that they be intrinsically public. Thus, even when such a value constitutes what a given agent most cares about—and so is most authentically his—nonetheless it remains a shared commodity, the common possession of all who are loyal to it. A public value is at large in the world, and while it is certainly made effective in the world through individuals, it transcends any of them.[11]

The Frankfurt-derived view of the will introduced in Chapter V and developed above provides further evidence against the notion that moral principles are effective—i.e., actually determine the will—through something like subsumption under a universally quantified proposition. I argued, following Frankfurt, that the will acts most authentically when it is determined to action by what it most cares about and the agent's actions reflect the internal structure of his will. As I noted in Chapter V, Frankfurt himself goes so far as to claim that he believes "we act morally when we are moved by love for a certain kind of world or a certain kind of life."

---

[11] Although it is true that most of my discussion of the developmental argument as concerned it application by parents to children, and apparently involves a smaller, more familiar sphere that I am discussing here, I think it is obvious that the kind of values that parents teach their children are public values of the sort I am here interested in. So when a child is taught not to take candy-bars at the grocery store he is being taught an instance of a much larger public value. To the extent that he successfully integrates a commitment of that sort in his will as a developing agent, then the public norm against stealing has effective force in his actions, and conditions the scope of possible actions in his practical deliberations. In some non-trivial way the values of the family are the values of the successful society, since, as Aristotle argued in the *Politics*, the family is both the first society and the foundation of all others.

In other words, morality has purchase on our souls through our love for and loyalty to it, not merely through some abstract notion of duty. When we are motivated by a concern for morality we are in fact expressing our love for and loyalty to "a certain kind of world or certain kind of life." Now the content of our understanding of that world—i.e., what makes it the world or life that it is—may seem to concern only universals. However, insofar as it is actually effective—insofar as morality actually motivates the agent—it does so because he cares about it; he *loves* it.

I think many moral philosophers have a tendency to think of morality as essentially involving a process of deliberation whereby we discharge universally quantified propositions to determinate, individually instantiated conclusions. The universally quantified propositions have a form similar to the following: (where 'x' ranges over moral agents and 'y' over actions) ∀x∃y (in conditions C, x should do y).[12] Thus, my acting morally consists in subsuming myself under this principle—i.e., instantiating myself as the 'x'. To act morally, then, is to *reason* properly, to recognize myself as a member of the set over which 'x' ranges, therefore to conclude that I *should* do y, and then actually to *do* y.

I do not want to claim it is impossible to reason and act thus. However, I think that such a coldly rational conception of morality is simply inaccurate in the vast majority cases. When our wills are actually determined to act morally it is seldom because we have arrived at the conclusion of a deductive argument of this sort. Rather, we become convinced (correctly or not) that

[12] This formula is meant to be very general, but not comprehensive—in particular, of course there may be more than one moral action (or perhaps none). It need not necessarily perspicaciously capture the form of every moral principle, nor does it rule out doing more than what morality demands. Rather, I only mean to suggest that moral principles tend to be understood under such a kind of universal quantification, *mutatis mutandis*.

something we care about is threatened and we are moved to act accordingly.[13] What we care about can and does vary widely, but even for the most sincere consequentialists and Kantians, I want to suggest when they actually act they generally do so because they are motivated by a kind of love.[14]

Whatever is the case with morality proper, the account of loyalty I have developed certainly works this way. When acting on a truly deep loyalty—i.e., when determined to act according to a Frankfurtian "volitional necessity"—the agent acts from his love for a certain value. Further, the sorts of values towards which loyalty aim (family, community, religion, etc.) tend to be public values integrated into the agent's will through his development in the way I outlined earlier. Accordingly, loyalty simply does not determine the will through judgments instantiating universally quantified propositions.

It might help to consider a particular case of loyalty—for instance, my loyalty to my mother. On the conventional model, if this loyalty has moral value then something like the following general proposition will be true: one ought to be loyal to his mother. This is the kind of proposition we examined in Chapter III and that Goodin and Pettit have in mind in their analysis of "duties of special concern"—it is a relativised duty whose "content is identified by back-reference… [with] a pronominal device, to the bearer." As before, this proposition can be formalized with a universal quantifier: ∀x (x ought to be loyal to x's mother). If I embrace such a proposition (i.e., hold it to

[13] Frankfurt's observation on this topic rings very true: "If someone tries to interfere with my children, I will try to fight him off. If someone tries to interfere with out efforts to bring about or to sustain the kind of world we love, why should we not be as confidently disposed to fight him off? We do not need any knockdown reason for knocking him down, other than the reason that he is trying to damage something we love" (Buss and Overton, 277).

[14] I know this point will be hotly contested; I shall have more to say about it below. What this claim essentially amounts to is a claim about how effective volition is regulated in by the structure of higher-order desires.

be true) and understand that I am a member of the set over which 'x' ranges, then I will conclude that I ought to be loyal to my mother. Arriving at such a conclusion, I might then take steps to make it effective in my actions. The normative force of my acting in accordance with the perceived duty here derives from the universally quantified proposition and my subsumption of myself under it. It does not derive from anything intrinsically particular about *this* woman or *my* will, except insofar as she stands in a certain formal (extrinsic) relationship to me.

If loyalties are best understood as a special class of universal moral duties, we would expect them to have a form similar to this (at least to the degree of being apt to formalization in some sort of universally quantified proposition).[15] However, the view of the will I have advanced throughout this dissertation denies that this is the right analysis of my loyalty to my mother. My loyalty to my mother grows out of my love for *her*, not my respect for a universally quantified principle under which I have subsumed myself. In other words, the normative force of my acting from the loyalty I owe her is a result of my will being oriented in a particular way—namely, in being partially constituted by my love for her. When I act out of such loyalty I give expression to the fact that her good is so deeply implicated in my will, that acting contrary to my perception of her good is simply (volitionally) impossible.

It is absolutely vital to see the distinction here: in the one case my loyalty to my mother is analyzed as an instantiation of an independently holding principle whose truth and moral force is wholly independent of my

[15] The scope of the universal quantifier may be limited in certain ways. For instance, the proposition that formalizes such a duty may go beyond pronominal back-reference to include, e.g., conjunctions which specify membership in certain groups, etc.

recognition of it. In the other case, my loyalty to my mother grows out of the place my love for her has in constituting my very will. Thus, in my view it is simply not the case that my loyalty to my mother grows out of a process of reasoning at all.

On this analysis my acting from loyalty does not represent the conclusion of an argument whose first premise is a universally quantified normative principle and whose second premise is my recognition of myself as a member of the set over which that quantifier ranges. Rather, it is an expression of the deepest structure of my will, which issues directly in action in consequence of a perceived threat to something (someone, in this case) I care deeply about. If she is in danger and I rush to save her my actions are probably not the result of any kind of deliberation at all, but the necessary consequence of having a will formed by and through my love for her. Thus, my commitment to the proposition "I ought to be loyal to my mother" is not merely an instantiation of my commitment to the general principle "everyone ought to be loyal to his mother." In terms of how my will expresses an action that embraces it, the normativity of the particular proposition is independent of that of the general one.[16]

Now it might be objected that I have just offered a kind of general principle (regarding the force of volitional necessity) in contradiction to my assertion that general principles are here inappropriate. This is not correct, however, because my point concerns the force of loyalty in determining how the agent actually acts. In other words, such a principle is descriptive of how

[16] This does not preclude the possibility that the general principle that everyone ought to be loyal to his mother is true as well, in fact I think it probably is. Rather the point is in how my loyalty effectively determines my will in such cases. If there is moral value in the universal statement it will be captured only in a will disposed to love the ideal it embodies.

and why the agent acts but is not normatively determinative of those actions. The central point is that an agent compelled to act by a volitional necessity (to save his beloved mother, for example) does not act on any principle whatsoever. Although I can describe the class of actions in terms of a principle, his agency is not founded on any sort of deliberation from principles. That is the key point; when acting on the deepest sort of loyalty an agent's actions are not motivated by the conclusion of a deliberative process in which the agent recognizes himself as properly subsumed under some universal moral principle.

This is true even for a moral theory that does not require perfect deliberation for every act (e.g., rule utilitarianism). Overall traditional moral theories require at least the justification of some standing intention or policy (rule, etc.) from a deliberative standpoint by applying a universal principle (e.g., pleasure maximization) to individual circumstances. My central claim is that the deepest loyalties in human life simply do not work that way; they make no essential reference to some set of universal principles. They are what they are simply because the agent loves what he loves. The "justification" of such loyalties lies not in their tendency to conduce to some favored state of affairs (or to describe the structure of practical reason, etc.) but in the fact that they are necessary to ground the very possibility of (moral) agency itself.

Now we can see is how misleading the analogy to learning arithmetic really is. In that case, the force (i.e., truth) of the particular instantiations (that these two oranges added to those two oranges, makes a total of four oranges) is entirely parasitic upon the truth of the general proposition "2 + 2 = 4."[17]

[17] Of course, I am speaking loosely here and not purporting to be making any claims about arithmetic. Perhaps, strictly speaking, the truth of any instantiation of arithmetic, including the adding of two oranges to two other oranges to yield a total of four oranges, is parasitic

Thus, once the child learns the principles of addition, he may put away the childish things of the oranges, pennies, or fingers. On the other hand, in the case of the deepest loyalties, the normative force of the particular lies not in its instantiation of a universal, but in its particular content—that I am loyal to my mother, *this* particular woman here, whose love for me became integrated into my very will as I developed to become the kind of creature who could love her back.

Thus, insofar as the agent embraces the values in which he was formed there can be no question of kicking the ladder away because the "ladder" becomes integrated within the will of the mature self. The central point here is that the normative force of loyalty that lies at the heart of agency is not derivative from any sort of universal principle independent of the volitional structure of the individual agent. In the case of arithmetic the particular modality whereby the student comes to understand the principles of addition are irrelevant to the force (i.e., truth) of those principles. The principle of addition would remain true in the absence of any particulars.

However, the force of loyalty (i.e., its role in bringing the agent to act) is intrinsically dependent on the particular object to which it is attached. This is simply because the normative force of loyalty is dependent on the structure of the will from which it proceeds, specifically the volitionally constitutive core commitments of the agent. What drives the agent to risk his life to rescue his beloved mother from danger has nothing to do with any principle whatsoever; it is how his will must move in reflection of its deepest constitution, directly manifesting in action what he most cares about.

---

only the nature of some set of numbers and the some property of that set or its members and not the proposition "2 + 2 = 4." Indeed this latter proposition might depend for its truth on the same facts. Nonetheless, I take it the point is clear enough.

The "ladder" cannot be discarded precisely because his mature will is itself constituted by the "rungs." That is, his will is itself partially constituted by his loves as formed through loyalty. These include his commitments to family, community, religion, etc. They are not dispensable precisely because their force is not parasitic on the truth of some more general principles, which the developmental process has revealed to him (though there may be general principles which describe these things that are true independent of him). Rather their normative *force* in the agent's life derives (even consists in) their particularity as loves and loyalties to individual, specific objects. Without those specific objects his love will become etiolated, his agency flattened, and perhaps even fatally undermined.

Consider for a moment that the object of my love for my mother is not merely whatever creature that happens to stand in some formal relationship to me, viz. the woman who gave birth to me. Rather, the object of my love for my mother is a particular, rather short, Asian woman that lives at such and such place and occasionally sends me Chinese cookies in the mail. It is true, of course, that she also satisfies that formal description, but my love for her is not directed at her under that formal description. I think that is clear enough if we consider that some other woman entirely might have stood under that formal description (e.g., a woman who gave me up for adoption, but whom I never knew), but that would not substantially change my volitional commitment to this particular short Asian woman (provided, of course, that our subsequent relationship was substantially the same, e.g. because she was the one who adopted me).

My deepest loyalties are directed at unique objects which form part of the content of my will through my deep love for them. I have literally come to

be, to have the will I have, on account of those loves that have formed in the developmental processes I have discussed above. Therefore, there can be no question of dispensing with these loves (and their particular objects), without dispensing with the very will, the very volitional structure, that informs the volitional necessities. As such, the normative *force* of my loyalties in bringing me to act lies intrinsically and inescapably in the particulars of my development, in the particular objects of my love and loyalty.

None of this is to say that the mature self cannot repudiate such values—of course, an agent can repudiate many of the values that go into constituting himself. However, in so doing he is consciously re-making himself, and whether or not that is good or noble depends on the particulars of both what he rejects and what he comes to affirm.[18] I want to stress, however, that the possibility of a deracinated agent who deliberately repudiates the values in which he was formed does not invalidate the analysis, for it can accommodate his "new" will as having much the same structure (i.e., constituted by deep, higher-order commitments) even though its content is very different.

One thing we certainly ought to note is that the deracinated agent is not more free or autonomous simply because he chooses to repudiate the values in which he was formed. For the agent who embraces and takes full responsibility for some of the public values which inform his society can do so just as autonomously as the one who repudiates those values.[19] For what is

---

[18] In other words, we cannot make the evaluation in purely formal terms. We would have to examine the particular public values of his society. Nonetheless, if I am right that there are real limits on what constitutes a "good enough" world—limits set by human nature—then *eo ipso* there will be some aspects of his society that are good, no matter how bad it is.

[19] Indeed, I would be inclined to think the agent who deliberately embraces some of his society's public values is acting *more* autonomously to the extent that his take responsibility

most important in this analysis is that the values the agent embraces be embraced as his own, freely integrated into his own will. Furthermore, even the most deracinated agent cannot fully extract himself from the public values of his society because they are encoded in the very possibility of sociability and social discourse (e.g., in language, the norms of social interaction, compromise, etc.).

This point re-emphasizes the importance of understanding development as a prerequisite for understanding the nature of moral maturity, for mature moral agency is founded upon a stable core of volitional commitments. The developmental model allows us to understand both how that volitional core arises in the first place and the nature of some of its most important content. Since the will is formed in dialectic with the public values of the loving and lovable social world the developing agent inhabits, it is the case that many of the social norms which inform that world will be written into his will through loyalty. This is so simply because the nature of their publicity requires it.

Just as there can be no private language, there can be no private morality. The intrinsic nature of moral commitments requires that they be shared within a moral community, just as the nature of language requires that it be shared among members of a speech community. Much as the developing speaker acquires within himself his speech community's linguistic structures, so the developing agent takes on his moral community's moral content into his will.

---

evidence of reflection, whereas repudiation is often merely tossed off. Of course, one can be reflective or unreflective in either case.

The analogy can be pressed even further, I think, in that just as the language one speaks informs in many ways how one thinks, so the moral world volitionally imbibed (so to speak) necessarily informs the possibility of one's moral practice and discourse. This is of course not to deny the possibility of a large overlap between different moral communities, just as there is among different speech communities. But as certain concepts and even patterns of thought are specific (and even unique) to speakers of particular languages, I think it is plausible to think that genuine moral content can be unique to members of particular moral communities.[20]

Of course propositions can be translated between languages and indeed there may be a "deep grammar" intrinsic to all human language. Just so moral content can be "translated" between moral communities, even if sometimes imperfectly. Furthermore, lest this be thought an invitation to simple relativism, this claim is fully consistent with there being a truth of the matter about morality. There may be (indeed I think there is) a "deep morality" which is constrained and determined by human nature. Nonetheless, there may be (and I think there are) different moralities just as there are different languages, and these may generate differing patterns of moral obligation.[21]

---

[20] Williams writes of such "content" as "thick" ethical concepts. These are concepts, particular to certain moral communities, which are intrinsically evaluative as well as descriptive. Their first-personal use requires membership in the communities in which they are current. (c.f., Williams [1985], 129, 142-5).

[21] Again, though, this does not lead to moral relativism if the "deep morality" which all moral communities share insofar as they share a constant human nature is ultimately determining. Thus, a society that has a social practice plainly in conflict with basic morality—e.g., suttee in 19th century India—can be and ought to be morally criticized. The grounds of that criticism will be precisely this "deep morality" and the mere fact of the communities deeply felt attachment to the practice will not, by itself, be sufficient to shield it from criticism. In a case like this it become necessary to judge that our confidence in the correctness of our moral condemnation outweighs any sort of sensitivity to the indigenous norms. Thus, as part of our commitment to morality we are required to effect a fundamental change in the social practices of the alien society and thus in the very nature of that moral community.

However this may be, real agents acting in the real world express their agency through wills formed in the context of particular social circumstances and have volitional commitments bearing the marks of those circumstances. An understanding of the nature of the development of the will through the dialectic of love mediated by loyalty is absolutely key to understanding both why individual moral communities take the shape they do and how individuals acting within those communities will express themselves through their practice of moral life and discourse. By the same token, an understanding of the nature of the developmental processes will help us to see the limits which human nature places on the scope of these possibilities.

The process of development I have discussed gives us clues towards these very limits. For if we can discover the necessary processes by which moral psyches come to be we will have a better idea of what it is that makes us human. Our researches have clearly pointed to the operations of a loving dialectic and the requirements of loyalty to the generation and inculcation of moral values. However, we need not stop at these broad limits. We can further specify the sorts of values that our particular moral community embodies and requires of its members.

Thus we require an investigation not only of the general process of moral maturation determined by the scope of human nature and common to all human societies. We need as well a further investigation into the particularities of our own moral community and thus the loyalties which inform the values we actually do have and express. Only then can we recognize that the moral values we express in the shared fabric of our moral community include not only those values of "deep morality" dictated by

human nature but also the particular and peculiar values that make our civilization, our society, our moral community, the unique one it is.

As I have repeatedly emphasized, loyalty stands at the heart of this, at the heart of the dialectic of love between the individual and his community. The nascent loyalty of the developing psyche drives the moral education of new members of the community. Mature loyalty ties the community together. Paternal loyalty makes possible the moral procreation of that community. Loyalty is thus *sine qua non* for any sort of moral life and thus for morality itself. Therefore, any moral theory or practice which militates against loyalty, militates against morality.

Contemporary moral theory can do precisely this to the extent that it makes us think that the ethical life must take a particular form involving justifications through deliberation from universals to particulars. Depending on how such universals are understood (e.g., Kant's notion of the pure law), they have the effect of radically disvaluing the particular and the local and thus undermining the foundation of a stable will. Thus, if I am correct in maintaining that all human agents must develop through the sort of dialectic described above, then the requirement that all actions be justified (even if only in principle) through moral reflection in terms of a universalist moral theory will have the effect of alienating agents from the only real moral content and context they can have.

This alienation has two particularly malign effects. First, it can yield a kind of rarefied indecision, an effete embrace of the complexity of the practical life such as to generate a kind of volitional paralysis in which the agent never

can quite bring himself to act.[22] By making the individual agent the bearer of all value, contemporary moral theories tend to require all genuinely ethical force to derive somehow from that individual's reflective capacities to reason out his duty (though particular individuals might act from fear of retribution, etc.). Such theories simply do not make enough space for the fact that much, even the vast majority, of ethical life is passed (quite well) in "deliberative silence." Indeed, it tends to go most smoothly (including generating the most ethically preferable actions) precisely when there is no need for deliberation.

The second malign effect is that this alienation divides morality from ethical life as a whole. It makes morality always a "problem," a challenge for the agent to fit morality into the other things he cares about. As such it makes the demands of morality come from "outside" in a way that makes them particularly disagreeable. In fact, the moral life should come from a disposition towards the virtues, and as such can be felt to come not only from the "inside" but also to be so natural that acting contrary to them can become the alien burden. If the moral virtues are successfully integrated into the agent's character through a successful moral maturation, then the ethical life will proceed from his own volitional constitution. Acting in accord with the demands of morality will itself become a kind of self-expressive loyalty. Thus, loyalty to the vision of a shared moral world is at the heart of a successful integration of morality into the lived ethical life.

---

[22] Williams writes, "One reason why conservatives and traditionalists attack reflection is that they fear the uncertainty that seems to follow from it, the situation in which the best lack all conviction. The result they fear is something to be feared, and they are right to detest a certain liberal posture that makes a virtue of uncertainty itself and, in place of conviction, enjoys the satisfactions—the equally intellectualist satisfactions—of refined indecision. But those traditionalist and those liberals share the error of thinking that what conviction in ethical life has to be is knowledge, that it must be a mode of certainty" (Williams [1985], 168-9). He goes on to argue for the necessity of having ethical "confidence" instead of certainty as I noted above.

In this chapter I have tried to show just how necessary it is to come to grips with the processes of moral development in order to actually understand what a morally mature will consists in. I have argued that it is only by taking account both of the internal structure of the will and the ways it comes to have that structure that we can have a plausibly compelling understanding of the ethical life as lived, as opposed to the artificial and abstract constructions of the moral theories. Finally, I have argued that loyalty is at the heart of these developmental processes and indeed at the heart of the ethical life itself. As such, I maintain that any tolerably complete philosophical understanding of morality must make an honored place for loyalty and that no moral theory that ignores the importance of loyalty can ever be adequate to the demands of the lived ethical life.

Conclusion

In the Introduction, I claimed that the proper subject matter of moral philosophy, above all else, is the constitution of a good will. My claim was that many sorts of moral theory, particularly consequentialism but the same is true in lesser degrees of others, simply seem not to be about the right sort of thing. This is because the force of any normative proposition is effective only insofar as it determines the will of the individual moral agent. Consequentialism makes the moral agent responsible for something over which he does not bear full and direct control—some state of affairs in the world as a whole. This seems to me to be simply a mistake, a misunderstanding of what moral philosophy is *about*.

What is ultimately true in any case is that the only real impact any moral theory has in the world is through the actions, habits, and practices of real human beings committed to that theory. Thus, even the moral philosopher must admit that no matter what account of the good he gives, it will be only as effective as real human beings order their wills in reference to that account. This is in one sense a trivial truth, but in another it is of the first importance. For it points to the fact that the *practice* of morality is very different from *thinking about* morality, a distinction to which I think too many moral philosophers fail to give adequate consideration.

Thus a true account of the nature and formal constitution of the will would cut across all lines of moral theory. That is, the correct account of the will has relevance no matter what the facts of the matter are concerning the deeply controversial truths of morality and the good. If, somehow, it really is the case—the fact of the matter—that the greatest happiness principle encapsulates the truth about *the* good, nonetheless it will only be effectively translated into moral practice through the wills of those committed to it.

Thus, an understanding of the nature and constitution of the will should be of interest to moral philosophers of all stripes. By the same token, since it is almost certainly true at minimum that *ought* implies *can*, the limits of the will most definitely represent the limits as well of moral philosophy. It is from reflections such as these that I am confident in claiming that even those moral philosophies which I have criticized must accommodate the nature of human wills as we actually have them, not merely as the adherents of those theories might like them to be.

For the moral intuitions which inform our valuing of human shaped wills as they actually are, while not indefeasible, nevertheless represent the core of ethics and the foundation of our moral life. While we may be able to abandon some of them in the pursuit of a coherent moral system, we should always be hesitant to give up any more of them than we might have to. My claim throughout this dissertation has been simply that the moral intuitions attached to loyalty are of this stronger sort, precisely because they lie at the very foundation of self-hood.

These considerations suggest that the investigation of the will is a necessary and important point of departure for moral philosophy, whatever the truth of the matter about the good turns out to be. It also suggests that the

account of the will in terms of loyalties can and does subsume all moral theories in the way suggested by Frankfurt. Because if morality is truly made effective by love through loyalty—the love, as he puts it, of "a certain kind of world or certain kind of life"—this is as fully true for the consequentialist, or Kantian, or moral contractualist, as it is for anybody else.

All of these theories of normative ethics are at the mercy of the truths of moral psychology. They can actually be effective only through the agents who order their lives around them. The sophisticated moral psychologist can then explain each contending theory of normative ethics entirely in terms of a loving commitment to "a certain kind of world or certain kind of life." Again, this is in one sense a trivial truth but in another it has very important implications.

For if the application of all normative ethics to life as it is really lived is reducible to the love of and loyalty to a certain kind of world or certain kind of life, then no normative ethics can make demands on the moral agent which undermine his capacity for having a will constituted by such a love. This is what stands at the heart of my project and my claims for loyalty. Loyalties—real, substantial, visceral, messy loyalties—alone make possible the kind of wills which can be constituted by love for a certain kind of world or certain kind of life. Morality itself is a kind of a loyalty—a loyalty to the good (even though what the good is understood to be can and does vary widely).[1]

[1] This I take to be an empirically true statement about the diversity of belief; it may nonetheless be the case (indeed I think it is) that there is a fact of the matter about the good. My point here is dialectical; whatever is true about the fact of the matter of the good, it nonetheless remains the case that humanly-shaped wills of the sort we value will necessarily be ordered by loyalties. Thus, whatever one's account of the good, effective morality—morality in a lived life—is a product of loyalty to that account, whether or not it is the correct account.

The combined tendency of moral philosophy to overemphasize rational deliberation and to undermine loyalty has the effect, in the end, of undermining itself. For without loyalties morality is empty, the plaything of effete moral theories, rather than, as it should be, the substance of a virtuous and humanly fulfilling life. Without an emphasis on, and recognition of, loyalty, morality and the moral life as a whole are turned into an aesthetic pastime.

I want to be clear: I am not merely saying that without genuine loyalty there can be no effective morality in the world (though I am saying that). I am going further to claim that without an understanding of the place loyalty occupies at the heart of rational agency itself, the practice of moral philosophy borders on the nonsensical. Thus, loyalty is not merely the quaint quasi-moral phenomenon it is often made out to be, the residue of a lack of moral reflection. It is the very substance and foundation of any sort of moral agency whatsoever. Until our moral theories take it seriously and accommodate themselves to its demands, they will not simply be inaccurate to the moral phenomenology of our lives (though they will be that), they will actually do violence to the very possibility of a full theory of normative ethics directed at the good for human beings.

In the Introduction, I quoted Bernard Williams' assertion that "unless such things [as deep attachments to other persons] exist, there will not be enough substance or conviction in a man's life to compel his allegiance to life itself. Life has to have substance if anything is to have sense, including adherence to the impartial system; but if it has substance, then it cannot grant supreme importance to the impartial system, and that system's hold on it will

be, at the limit, insecure." I have suggested that this entire dissertation is a kind of marginalia on this notion of a life having "substance." I want now to draw together a few observations and conclusions along these lines.

The analysis of the will that I have taken over from Frankfurt and modified with the developmental account is precisely intended to help us understand what Williams' notion of a life with substance actually entails. The substance of a life is its specifying content—what makes it the particular life it is and what makes any given agent the particular agent he is. The substance of a life is found in the loves and loyalties that the agent has, because these loves are the very foundation stones from which his rational agency arises.

The crucial mistake that Kant makes is in thinking that there can be a kind of formal principle of causality (the pure will) which alone provides the foundation for all moral responsibility and moral agency in a transcendental ego. However, I think many more philosophers are guilty of an analogous mistake, taking practical *reasoning* (under various descriptions) to be the only foundation of rational agency. It is primarily because their analyses are less rigorous than Kant's that this is less immediately apparent. Nonetheless, even consequentialists, moral contractualists, and many other sorts of moral theorists, fall into a similar trap.

All these various theories treat the moral agent primarily as a rational decision maker. This turns morality in general, and morally responsible agency in particular, into primarily a process of making rational (maximizing) judgments. For sure, the various theories differ wildly on the good to be maximized (e.g., pleasure, "negative freedoms", "social justice", etc.), but they share a commitment to arriving at a view that will rationally compel the moral agent to act morally. What matters in any given agent is not his particularity at

all, but rather some property (generally, but not always, rationality) that he shares with all other moral agents and to which he has no more exclusive title or claim than anyone else.[2]

Yet the paradox is that this rationality, which each of the contending parties claims to best represent, is in itself empty. For the actual human creatures for whom morality matters, who instantiate the only truly existing tokens of the universal rationality, are, *qua* moral creatures, much more than their respective capacities for rationality. Indeed, their capacity to reason is itself dependent on their logically and temporally prior existence as (moral) selves, which is in turn dependent on their having wills constituted and formed by love through loyalty.

For purposes of moral psychology I think that the rationality of the agent is of much less importance than is generally acknowledged. I certainly do not mean to deny that agents ought to be rational, but their true moral worth and significance lies in something much more than their capacity for rational choice. For choice itself, as Frankfurt repeatedly insists, is founded on necessity. It is only because my choices issue from a will that is itself stable that they count as choices at all. Furthermore, as I have been at pains to argue, the stable self that is at the very bottom of all agency is itself and must be unchosen, for its formation is necessarily a dialectic of love in which it is quite literally called forth by that love. Thus, the very cornerstone of the rational agent's agency lies not in anonymous reason, but in his particular personality,

[2] Of course a consequentialist might focus on some property like feeling pleasure, and so, like Peter Singer for example, argue that the class of relevant creatures whose interests must be consulted should be expanded beyond those possessing rationality. Even if this is so, however, the point is that the moral imperative to act on such judgments is itself advanced as a requirement of rationality (in this case, the exclusion of purportedly rationally unjustified discriminations against the lower animals).

the personality he owns because of what and whom he loves, and which has the shape that it does because of what and who loved him.

This, then, is the substance of a human life, which is at the heart of allegiance to life itself. For it *is* life itself—life is a product of love and loyalty in every sense. Maternal and parental love brings forth the new psyche, loyalty to the public values of society give it content, and acting on loyalty to those values constituting the soul embodies moral agency itself. These values—the agent's loves and loyalties—literally define the content of the only thing which can truly be identified as the source of action, as the locus of causation, the ground of agency. Rationality, without love and volition, is empty. Loyalty is simply and unavoidably the consequence of having a human will shaped by and expressive of love.

It should be obvious then that this substance must be particular, for loves are particular. Loves are not fungible, even when new love comes it does not fully replace the old. It is an imbecile who wonders why a new child does not alleviate the loss of his older sibling in the heart of his mother, or even, less importantly but just as truly, that the new pet cannot simply take the place of the old. For love to be love it must be particular; it must have a unique object. It is these sorts of particularities which underwrite the deepest loyalties.

So loyalty becomes central to the will, because loyalties are inescapable expressions of the particular loves which constitute a human shaped will. Thus loyalty is truly at the heart of Williams' notion of substance. A soul with substance—a human shaped soul, a soul that bears and embodies human value—simply is a *loyal* soul. Loyalty simply is the expression of substance. I hope that now we can finally begin to appreciate the enthusiasm (which was

perhaps somewhat more surprising when we first saw it in Chapter II) with which Royce proclaimed: "Loyalty… says to us, 'In this cause is your life, your will, your opportunity, your fulfillment.'"

A Formal Definition

With these observations in hand we can now offer a formal definition of the deepest sorts of loyalty. There will be no surprises here, but nonetheless it is useful to make a clear statement. *At the limit, loyalty is a property of the will, expressive of the will's internal structure as a hierarchy of desires formed in response to and as a reflection of a loving world.*

The parts of this definition of course partake of the nature of the analysis as I have tried to advance it through Part II of this dissertation. Again, I reemphasize that loyalty applies to the will, but as I suggested in the Introduction, it is not merely an instrumental property like intelligence or courage whose value is entirely dependent on the object of loyalty. For, as I have tried to show in my the account of the development of the moral psyche, loyalties are intrinsic both to the very advent and constitution of the soul. We can say without reservation that without loyalty there would simply be no human souls.

The upshot of this is that the expression of loyalty is simply intrinsic to the human condition. Loyalties are not merely inevitable but participate at the deepest level of rational agency, because it is through loyalties that the will itself is constituted, that the personality and identity of the agent is formed. Furthermore, because human souls come to be in response to, and are

informed by, the love they receive from the world around them, loyalties stand at the very heart of what it is to be human.

This is why so much of my concern has revolved around the admittedly nebulous notion of a *human-shaped* will. It has certainly never been my intention to define strictly the exact dimensions of possible human wills. My project has been much less ambitious than that, but at the same time I think rather more plausible. My argument has proceeded by claiming to better explain the phenomenology of loyalty, of both its felt importance in human life, and its pervasiveness in human society. But as well, my argument holds that the phenomenology points us towards these deeper truths about the true nature of agency and the soul.

If I am correct in even the broadest terms, loyalty of some sort participates in the formation of every human being, and in living any sort of valuable life at all every human being expresses loyalty to something beyond himself. However, I have also had a more specific goal, to reconcile this truth to the practice of moral philosophy in our contemporary Western tradition. For I think that once we recognize how central loyalties are to the development of the moral self we must come to recognize just how important loyalties are in maintaining the possibility of the practice of a moral life.

Once we realize that loyalties, and the public values which they embody, are intrinsic to the very formation of the souls of the young and the maintenance of the souls of the morally mature, we ought to recognize that they should and must be embraced. We should come to realize that it is a *moral* imperative to defend and protect the values lying behind the loyalties to our community and its traditions, in which the very existence of that moral community as a moral community are at stake. For if it is the case that without

loyalties there can be no moral selves, then it is certainly the case that without a vigorous public defense and inculcation of loyalty a moral community slowly commits suicide. This does not mean that any moral community should be free from criticism, but it does mean that those features of the moral community that make moral life possible (e.g., commitments to the truth and some sort of mere morality, etc.) must be affirmed through public loyalty.

The formal definition I have offered makes very clear that loyalty mediates between the individual and society. Loyalties are the very means by which individuals are tied together to each other in moral community. This is because the very effectiveness and practice of morality depends most not on the individual's capacity for moral reflection but on his deep seated and almost pre-reflective propensities to express the public values in which he was formed. We must move beyond conceiving of morality as practical reasoning for the simple and obvious reason that real moral agents, be they philosophers or the man on the street, generally and for the most part actually practice morality not because of moral reflection, but ingrained social prejudice. And of course this social prejudice simply is a kind of loyalty to the community as a whole, its constituent members, and its traditions.

I would now like to close by acknowledging the epigraph that surmounts the head of this dissertation, several hundred pages back. There I offer a quote from Josiah Royce: "When a man feels his present ties to be arbitrary or to be a mechanical bondage, he sometimes says that it is irrational to be a mere spoke in a wheel. Now, a loyal self, is always more that a spoke in a wheel. But still, at worst, it is better to be a spoke in a wheel that a spoke out of the wheel." The sort of prejudice Royce captures here is at the heart of the

"case" against loyalty.

Loyalties are often taken to be arbitrary or mechanical; they are unchosen, so how can they be morally significant? They are the product of prejudice and habit, not of moral reflection and practical reason; so how can they be justified? Yet I hope the reader will now be more ready to consider with sympathy the possibility that without these things we simply can have no morality, not just no moral philosophy, but no *morality*. Since moral philosophy serves morality, not vice versa, it must bow before the demands of moral psychology, the demands of the lived moral life.

If my account of developmental moral psychology is anywhere close to correct, we are all spokes in wheels not of our own making. We might like to think that this is somehow demeaning to our moral autonomy, but in fact we have no other choice. If we do not like being spokes in whichever wheel we find ourselves, the only other option is to be spokes in another one, or worst of all to be out place altogether. We are spokes whether we see it or not, and as moral philosophers we must make our peace with this if we are to undertake the task of pursuing moral knowledge.

BIBLIOGRAPHY

Aristotle. *The Nicomachean Ethics*, translated by T. Irwin (Indianapolis: Hackett Pub., 1999).

Augustine. *Confessions*. Translated by Henry Chadwick. (Oxford: Oxford UP, 1991).

_____. *On Free Choice of the Will*. Translated by Thomas Williams. (Indianapolis: Hackett Pub.,1993).

Baron, Marcia. "Impartiality and Friendship" in in *Ethics*, Vol. 101: Issue 4 (Jul. 1991), pp. 836-857.

Bratman, Michael. "Reflection, Planning, and Temporally Extended Agency" in *The Philosophical Review*, Vol. 109: No. 1 (January 2000), pp 35-62.

Burke, Edmund. *Edmund Burke: Selected Writings and Speeches*. Edited by Peter J. Stanlis. (Washington, D.C. : Regnery Publishing, 1963).

_____. *Select Works of Edmund Burke*: a new imprint of the Payne edition / foreword and biographical not by Francis Canavan in 3 vols. (Indianapolis: Liberty Fund, 1999).

Buss, Sarah and Overton, Lee, editors. *Contours of Agency: Essays on Themes fromHarry Frankfurt*. (Cambridge, MA: MIT Press, 2002).

Castell, Alburey, editor. *Essays in Pragmatism* (New York: Hafner Press, 1948). *in which, especially*:
James, William, "The Will to Believe" pp. 88-109.

Dworkin, Ronald. "Foundations of Liberal Equality" in *Equal Freedom*, edited by Stephen Darwall. (Ann Arbor, MI: Univ. of Michigan Press, 1995). pp. 190-206.

_____. *Taking Rights Seriously*. (Cambridge, MA: Harvard UP, 1977).

Fletcher, George. *Loyalty: An Essay on the Morality of Relationships*. (New York: Oxford UP, 1993).

Foot, Philippa. *Virtues and Vices and Other Essays in Moral Philosophy*. (Oxford: Oxford UP, 2002).

Förster, Eckart, editor. *Kant's Transcendental Deductions*. (Stanford, CA: Stanford UP, 1989).

Frankfurt, Harry. *The Importance of What We Care About: Philosophical Essays*. (Cambridge: Cambridge UP, 1988).
in which, especially:
"Freedom of the will and the concept of a person"
"The importance of what we care about"
"Identification and wholeheartedness"

_____. "Duty and Love" in *Philosophical Explorations*, Nr. 1, January 1998.

_____. *Necessity, Volition, and Love*. (Cambridge: Cambridge UP, 1999).
in which, especially:
"On the Usefulness of Final Ends"
"On the Necessity of Ideals"
"Autonomy, Necessity, and Love"
"On Caring"

_____. "The Dear Self" in *Philosophers' Imprint*, Vol. 1: No. 1, Jan. 2001.

_____. "Some Mysteries of Love," the 2001 Lindley Lecture at The University of Kansas.

_____. *The Reasons of Love*. (Princeton, NJ: Princeton UP, 2004).

Friedman, Marilyn. "The Practice of Partiality" in *Ethics*, Vol. 101: Issue 4 (Jul. 1991), pp. 818-835.

Gauthier, David. *Morals by Agreement*. (Oxford: Oxford UP, 1986).

Goodin, Robert. "What is so Special About our Fellow Countrymen?" in *Ethics*, Vol. 98: Issue 4 (Jul. 1988), pp. 663-686.

Goodin, Robert and Pettit, Philip. "The Possibility of Special Duties" in *The Canadian Journal of Philosophy*, Vol. 16: No. 4, Dec. 1986, pp. 651-676

Hegel, G. W. F. *Elements of the Philosophy of the Right*. Translated by H. B. Nisbet, edited by Allen W. Wood. (Cambridge: Cambridge UP, 1991).

_____. *Phenomenology of Spirit*. Translated by A.V. Miller. (Oxford: Oxford UP, 1977).

Herman, Barbara. "Agency, Attachment, and Difference" in in *Ethics*, Vol. 101: Issue 4 (Jul. 1991), pp. 775-797.

_____. *The Practice of Moral Judgment.* (Cambridge, MA: Harvard UP, 1993).

_____. "The Scope of Moral Requirement" in *Philosophy & Public Affairs*, Vol.

30: No. 3 (2002), pp. 227-256.

Hume, David. *Essays: Moral, Political, and Literary*, revised ed. Edited by Eugene F. Miller. (Indianapolis: Liberty Fund, 1985).

_____. *Treatise on Human Nature*, 2nd edition. Edited by L.A. Selby-Bigge and P.H. Nidditch. (Oxford: Oxford UP, 1978).

Kant, Immanuel. *Critique of Practical Reason*. Translated by Lewis White Beck. (New York: Macmillan, 1993).

_____.*Critique of Pure Reason*. Translated by Paul Guyer and Allen Wood. (Cambridge: Cambridge UP, 1998).

_____. *Groundwork of the Metaphysics of Morals*, 3rd ed. Translated by James W. Ellington. (Indianapolis: Hackett Pub., 1981).

_____. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. (Cambridge: Cambridge UP, 1997).

_____. *Lectures on Ethics*. Translated by Louis Infield. (Indianapolis: Hackett Pub.,1980).

_____. *Metaphysics of Morals*. Translated and edited by Mary Gregor. (Cambridge: Cambridge UP, 1996).

Kirk, Russell. *The Roots of American Order*. (La Salle, IL: Open Court, 1977).

Kojève, Alexandre. *Introduction to the Reading of Hegel*. Assembled by Raymond Queneau, edited by Allan Bloom, translated by James H. Nichols, Jr. (Ithaca, NY: Cornell UP, 1980).

Klein, Melanie. *The Psychoanalysis of Children*. translated by Alix Strachey. (New York: Grove Press, Inc., 1960).

Kymlicka, Will. *Multicultural Citizenship*. (Oxford: Oxford UP, 1995).

Lear, Jonathan. *Love and Its Place in Nature*. (New York: Farrar, Straus, and Giroux, 1990).

MacIntyre, Alasdair. *After Virtue: A Study in Moral Theory*, 2nd ed. (Notre Dame, IN: Notre Dame UP, 1984a).

_____. "The Magic in the Pronoun 'My'" in Ethics, Vol. 94: Issue 1 (Oct. 1983), pp. 113-125.

MacIntyre, Alasdair. "Is Patriotism a Virtue?" Published as a pamphlet by the

University of Kansas, 1984b.

_____. *Whose Justice? Which Rationality?* (Notre Dame, IN: Notre Dame UP, 1988).

Miller, David. *On Nationality*. (Oxford: Oxford UP, 1995).

Miller, Richard W. "Ways of Moral Learning" in *The Philosophical Review*, Vol. 94: Issue 4 (Oct., 1985), pp. 507-556.

_____. "Moral Contractualism and Moral Sensitivity: Critique and Reconstrual" in *Social Theory and Practice*, Vol. 28, No. 2 (Apr. 2002).

Nozick, Robert. *Anarchy, State, and Utopia*. (New York: Basic Books, 1974).

Oakeshott, Michael. *Rationalism in Politics and Other Essays*, new and expanded ed. (Indianapolis: Liberty Fund, 1991).

Oldenquist, Andrew. "Loyalties" in *The Journal of Philosophy*, Vol. 79: Issue 4 (1982), pp. 173-193.

_____. *The Non-Suicidal Society*. (Bloomington, IN: Indiana UP, 1986).

O'Neill, Onora. *Constructions of Reason*. (Cambridge: Cambridge UP, 1989).

Plato. *Plato: Complete Works*. edited by John M. Cooper (Indianapolis: Hackett Pub., 1997).
in which, particularly:
The Apology, translated by G.M.A. Grube.
The Crito, translated by G.M.A. Grube.
The Euthyphro, translated by G.M.A. Grube.
The Phaedrus, translated by Alexander Nehamas and Paul Woldruff.
The Republic, translated by G.M.A. Grube, rev. C.D.C. Reeve.

Railton, Peter. "Alienation, Consequentialism, and the Demands of Morality" in *Philosophy and Public Affairs*, Vol. 13, Issue 2 (Spring, 1984), 134-171.

Rawls, John. *A Theory of Justice*, revised ed. (Cambridge, MA: Harvard UP, 1999).

_____. *Political Liberalism*. (New York: Columbia UP, 1993).

_____. *Law of Peoples*. (Cambridge, MA: Harvard UP, 1999).

Raz, Joseph. "Liberating Duties" in *Law and Philosophy*, Vol. 8 (1989) pp. 3-21.

_____. *The Morality of Freedom*. (Oxford: Oxford UP, 1986).

Rousseau, Jean-Jacques. *The Basic Political Writings*. Translated by Donald A. Cress. (Indianapolis: Hackett Pub., 1987).
in which, especially:
"Discourse on the Origin of Inequality"
"On the Social Contract"

Royce, Josiah. *The Philosophy of Loyalty* in The Vanderbilt Library of American Philosophy Series. (Nashville, TN: Vanderbilt University Press, 1995). (Originally published 1908 by Macmillan Co.).

Sandel, Michael. *Democracy's Discontents: America in Search of a Public Philosophy*. (Cambridge, MA: Harvard UP, 1996).

Scanlon, T. M. *What We Owe to Each Other*. (Cambridge, MA: Harvard UP, 1998).

Scheffler, Samuel. *The Rejection of Consequentialism*, revised edition. (Oxford: OUP, 1982).

_____. Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought. (Oxford: OUP, 2001).

Scruton, Roger. *The Meaning of Conservatism*, 2nd ed. (London: Macmillan, 1984).

_____. *Philosopher on Dover Beach*. (New York: St. Martin's Press, 1990).

Smart, J.J.C. and Williams, Bernard. *Utilitarianism For & Against*. (Cambridge: Cambridge UP, 1973).

Stocker, Michael. "The Schizophrenia of Modern Ethical Theories" in *The Journal of Philosophy*, Vol. 73: Issue 14 (1976), pp. 453-466.

Taylor, Charles. *Human Agency and Language: Philosophical Papers 1*. (Cambridge: Cambridge UP, 1985).

_____. *Philosophy and the Human Sciences: Philosophical Papers 2*. (Cambridge: Cambridge UP, 1985).

_____. *Sources of Self*. (Cambridge, MA: Harvard UP, 1989).

Velleman, J. David. "Love as Moral Emotion" in *Ethics*, Vol. 109, No. 2 (Jan., 1999), 338-374.

Watson, Gary. "Free Agency" in *The Journal of Philosophy*, Vol. 72: Issue 8 (1975), pp. 205-220.

Williams, Bernard. *Ethics and the Limits of Philosophy*. (Cambridge, MA: Harvard UP, 1985).

_____. *Making Sense of Humanity*. (Cambridge: Cambridge UP, 1995).

_____. *Moral Luck*. (Cambridge: Cambridge UP, 1982).

_____. *Morality*. (Cambridge: Cambridge UP, 1972).

Wood, Allen W. *Kant's Ethical Thought*. (Cambridge: Cambridge UP, 1999).